

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 17:58:49 ; Search time 52 Seconds

(without alignments)
532.844 Million cell updates/sec

Title: US-10-071-976-2

Perfect score: 1270
Sequence: 1 MSKGEELFTAVVPIVLELDG.....VLEFYVTAAGTTHGMDELK 238

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 11641973 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 0%

Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubppaa/US07_NEM_PUB.pep:*
2: /cgn2_6/ptodata/1/pubppaa/PCR_NEM_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEM_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/CTUS_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/US08_NEM_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_NEM_PUB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US10_NEM_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US60_NEM_PUB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1270	100.0	238	9	US-09-575-847-2
2	1270	100.0	238	9	US-10-071-976-2
3	1246	98.1	238	9	US-10-057-505-2
4	1246	98.1	238	10	US-09-884-681-2
5	1246	98.1	238	12	US-10-024-686-2
6	1242	97.8	238	9	US-09-866-538-2
7	1242	97.8	238	9	US-09-900-345A-125
8	1242	97.8	238	9	US-10-121-258-10
9	1242	97.8	238	9	US-10-221-461-6
10	1242	97.8	238	10	US-09-920-922-4
11	1242	97.8	238	10	US-09-852-000-1
12	1239	97.6	243	9	US-09-900-345A-60
13	1239	97.6	243	9	US-09-900-345A-62
14	1239	97.6	243	9	US-09-900-345A-64
15	1239	97.6	243	9	US-09-900-345A-66
16	1239	97.6	243	9	US-09-900-345A-68
17	1239	97.6	243	9	US-09-900-345A-70
18	1239	97.6	1070	9	US-10-001-486B-2
19	1239	97.6	1099	9	US-10-259-864-4

20	1239	97.6	1147	9	US-10-259-864-1	Sequence 1, Appli
21	1239	97.6	1452	12	US-10-050-673-2	Sequence 2, Appli
22	1238	97.5	243	9	US-09-900-345A-54	Sequence 54, Appli
23	1238	97.5	243	9	US-09-900-345A-56	Sequence 56, Appli
24	1238	97.5	243	9	US-09-900-345A-58	Sequence 58, Appli
25	1238	97.5	243	9	US-09-900-345A-114	Sequence 114, App
26	1238	97.5	243	9	US-09-900-345A-116	Sequence 116, App
27	1238	97.5	243	9	US-09-900-345A-118	Sequence 118, App
28	1238	97.5	243	9	US-09-900-345A-120	Sequence 120, App
29	1237	97.4	243	9	US-09-900-345A-2	Sequence 2, Appli
30	1237	97.4	243	9	US-09-900-345A-4	Sequence 4, Appli
31	1237	97.4	243	9	US-09-900-345A-6	Sequence 6, Appli
32	1237	97.4	243	9	US-09-900-345A-8	Sequence 8, Appli
33	1237	97.4	243	9	US-09-900-345A-10	Sequence 10, Appli
34	1237	97.4	243	9	US-09-900-345A-12	Sequence 12, Appli
35	1237	97.4	243	9	US-09-900-345A-14	Sequence 14, Appli
36	1237	97.4	243	9	US-09-900-345A-16	Sequence 16, Appli
37	1237	97.4	243	9	US-09-900-345A-18	Sequence 18, Appli
38	1237	97.4	243	9	US-09-900-345A-20	Sequence 20, Appli
39	1237	97.4	243	9	US-09-900-345A-22	Sequence 22, Appli
40	1237	97.4	243	9	US-09-900-345A-24	Sequence 24, Appli
41	1237	97.4	243	9	US-09-900-345A-26	Sequence 26, Appli
42	1237	97.4	243	9	US-09-900-345A-28	Sequence 28, Appli
43	1237	97.4	243	9	US-09-900-345A-30	Sequence 30, Appli
44	1237	97.4	243	9	US-09-900-345A-32	Sequence 32, Appli
45	1237	97.4	243	9	US-09-900-345A-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1
US-09-575-847-2
; Sequence 2, Application US/09575847
; Publication No. US20030013149A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: WACHTER, Rebekka
; APPLICANT: REMINGTON, James
; TITLE OF INVENTION: LONG WAVELENGTH ENGINEERED FLUORESCENT PROTEINS
; FILE REFERENCE: RESEN1250-5
; CURRENT APPLICATION NUMBER: US/09/575, 847
; CURRENT FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 08/974,737
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: US 08/911,825
; PRIOR FILING DATE: 1997-08-15
; PRIOR APPLICATION NUMBER: US 08/706,408
; PRIOR FILING DATE: 1996-08-30
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea
US-09-575-847-2

Query Match 100.0%; Score 1270; DB 9; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.3e-113;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MSKGEELTAVVPIVLELDGNGHFKFSVSGEGDVTYTKLTKFICTTGKLPVMPPTL 60
1 MSKGEELTAVVPIVLELDGNGHFKFSVSGEGDVTYTKLTKFICTTGKLPVMPPTL 60
61 VTTFESYGVQCFSRYPDHHKRRHDFEKSAMPEGVQORTLFFKDDGNYKTRAEEKGGDTLV 120
61 VTTFESYGVQCFSRYPDHHKRRHDFEKSAMPEGVQORTLFFKDDGNYKTRAEEKGGDTLV 120
121 NRIELKGIIDFKEDGNIILHKLLEYKNNYINMADKONGIKVNFKIRHNIEDSGVOLAD 180
121 NRIELKGIIDFKEDGNIILHKLLEYKNNYINMADKONGIKVNFKIRHNIEDSGVOLAD 180
121 NRIELKGIIDFKEDGNIILHKLLEYKNNYINMADKONGIKVNFKIRHNIEDSGVOLAD 180

QY 181 YVOONTPLDGPVLLPDNHYLSTOSALS KDPNEKRDMHVLLEFYTAGITGHMDELYK 238
Db 181 YVOONTPLDGPVLLPDNHYLSTOSALS KDPNEKRDMHVLLEFYTAGITGHMDELYK 238

RESULT 2

US-10-071-976-2
; Sequence 2, Application US/10071976
; Publication No. US20030036178A1
; GENERAL INFORMATION:
; APPLICANT: Tsien et al., Roger Y. et al.
; TITLE OF INVENTION: LONG WAVELENGTH MUTANT FLUORESCENT
; PROTEINS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/071,976
; FILING DATE: 05-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/465,142
; FILING DATE: 1999-DEC-16
; APPLICATION NUMBER: 08/974,737
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07257/056001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-071-976-2

Query Match 100.0%; Score 1270; DB 9; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.3e-113;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVLELDGVDNGHKFSVSGEGSDVTYGLTKLFICTTGKLPVMPPTL 60
Db 1 MSKGEELFTAVVPLVLELDGVDNGHKFSVSGEGSDVTYGLTKLFICTTGKLPVMPPTL 60
QY 61 VTTFSSVGOCFSRPDMKRRHDFPKSAMPEGYVOORTIFFKDDGNYKTRAEVFEQDTLV 120
Db 61 VTTFSSVGOCFSRPDMKRRHDFPKSAMPEGYVOORTIFFKDDGNYKTRAEVFEQDTLV 120
QY 121 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 181 YVOONTPLDGPVLLPDNHYLSTOSALS KDPNEKRDMHVLLEFYTAGITGHMDELYK 238
Db 181 YVOONTPLDGPVLLPDNHYLSTOSALS KDPNEKRDMHVLLEFYTAGITGHMDELYK 238

RESULT 3

US-10-057-505-2
; Sequence 2, Application US/10057505
; Patent No. US20020164674A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: AURORA BIOSCIENCES CORPORATION
; APPLICANT: TSJEN, Roger
; APPLICANT: HEIM, Roger
; APPLICANT: CUBITT, Andrew
; TITLE OF INVENTION: TANDEM FLUORESCENT PROTEIN CONSTRUCTS
; FILE REFERENCE: REGEN1260-3
; CURRENT APPLICATION NUMBER: US/10/057,505
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: US 08/792,553
; PRIOR FILING DATE: 1997-01-31
; PRIOR APPLICATION NUMBER: US 09/396,003
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-057-505-2

Query Match 98.1%; Score 1246; DB 9; Length 238;
Best Local Similarity 97.9%; Pred. No. 2.5e-111;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVLELDGVDNGHKFSVSGEGSDVTYGLTKLFICTTGKLPVMPPTL 60
Db 1 MSKGEELFTAVVPLVLELDGVDNGHKFSVSGEGSDVTYGLTKLFICTTGKLPVMPPTL 60
QY 61 VTTFSSVGOCFSRPDMKRRHDFPKSAMPEGYVOORTIFFKDDGNYKTRAEVFEQDTLV 120
Db 61 VTTFSSVGOCFSRPDMKRRHDFPKSAMPEGYVOORTIFFKDDGNYKTRAEVFEQDTLV 120
QY 121 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 181 YVOONTPLDGPVLLPDNHYLSTOSALS KDPNEKRDMHVLLEFYTAGITGHMDELYK 238
Db 181 YVOONTPLDGPVLLPDNHYLSTOSALS KDPNEKRDMHVLLEFYTAGITGHMDELYK 238

RESULT 4

US-09-884-681-2
; Sequence 2, Application US/09884681
; Patent No. US20020061546A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Cubitt, Andrew B.
; TITLE OF INVENTION: Assays for Protein Kinases Using
; Fluorescent Protein Substrates
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/884,681
; FILING DATE: 19-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:

QY 181 YGQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238
Db 181 HYQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238

RESULT 7

US-09-900-345A-125
; Sequence 125, Application US/09900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; FILE REFERENCE: 10338-505
; CURRENT APPLICATION NUMBER: US/09/900.345A
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU P8078
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/00008
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP humanized
US-09-900-345A-125

Query Match 97.8%; Score 1242; DB 9; Length 238;
Best Local Similarity 97.5%; Pred. No. 6.1e-111;
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVPLVPLVELDGDVNGHKSFSVSGEGSDVYTKLTKFICTTGKLPVMPPTL 60
Db 1 MSKGEELFTAVPLVPLVELDGDVNGHKSFSVSGEGSDVYTKLTKFICTTGKLPVMPPTL 60
QY 61 VTTFSSYGVCFSRYPDMKRRHDFKSAPEGYVOORTIFFKDDGNKTRAEVKFEGDPTLV 120
Db 61 VTTFSSYGVCFSRYPDMKRRHDFKSAPEGYVOORTIFFKDDGNKTRAEVKFEGDPTLV 120
QY 121 NRLEKGIIDKEDGNILGHKLENYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRLEKGIIDKEDGNILGHKLENYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 181 YGQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238
Db 181 HYQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238

RESULT 8

US-10-121-258-10
; Sequence 10, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; FILE REFERENCE: UC083.1CP2CPI
; CURRENT APPLICATION NUMBER: US/10/121.258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 238

; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-121-258-10

Query Match 97.8%; Score 1242; DB 9; Length 238;
Best Local Similarity 97.5%; Pred. No. 6.1e-111;
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVPLVPLVELDGDVNGHKSFSVSGEGSDVYTKLTKFICTTGKLPVMPPTL 60
Db 1 MSKGEELFTAVPLVPLVELDGDVNGHKSFSVSGEGSDVYTKLTKFICTTGKLPVMPPTL 60
QY 61 VTTFSSYGVCFSRYPDMKRRHDFKSAPEGYVOORTIFFKDDGNKTRAEVKFEGDPTLV 120
Db 61 VTTFSSYGVCFSRYPDMKRRHDFKSAPEGYVOORTIFFKDDGNKTRAEVKFEGDPTLV 120
QY 121 NRLEKGIIDKEDGNILGHKLENYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRLEKGIIDKEDGNILGHKLENYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 181 YGQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238
Db 181 HYQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238

RESULT 9

US-10-221-461-6
; Sequence 6, Application US/10221461
; Publication No. US20030092902A1
; GENERAL INFORMATION:
; APPLICANT: Marsh, Donald J.
; TITLE OF INVENTION: MELANIN CONCENTRATING HORMONE RECEPTOR
; FILE REFERENCE: 20652P
; CURRENT APPLICATION NUMBER: US/10/221.461
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: PCT/US01/08071
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/189,698
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-221-461-6

Query Match 97.8%; Score 1242; DB 9; Length 238;
Best Local Similarity 97.5%; Pred. No. 6.1e-111;
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVPLVPLVELDGDVNGHKSFSVSGEGSDVYTKLTKFICTTGKLPVMPPTL 60
Db 1 MSKGEELFTAVPLVPLVELDGDVNGHKSFSVSGEGSDVYTKLTKFICTTGKLPVMPPTL 60
QY 61 VTTFSSYGVCFSRYPDMKRRHDFKSAPEGYVOORTIFFKDDGNKTRAEVKFEGDPTLV 120
Db 61 VTTFSSYGVCFSRYPDMKRRHDFKSAPEGYVOORTIFFKDDGNKTRAEVKFEGDPTLV 120
QY 121 NRLEKGIIDKEDGNILGHKLENYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRLEKGIIDKEDGNILGHKLENYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 181 YGQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238
Db 181 HYQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238

RESULT 10

US-09-920-922-4
; Sequence 4, Application US/09920922
; Patent No. US20020083488A1


```

; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-4

Query Match          97.8%; Score 1242; DB 10; Length 238;
Best Local Similarity 97.5%; Pred. No. 6,1e-111;
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGELFTAVVPIVLVDGVDVNGHKFVSVSGEGEDVYTKLTKFKICTTGKLPVWPPTL 60
DB 1 MSKGELFTGVVPIVLVDGVDVNGHKFVSVSGEGEDATYKTLTKFKICTTGKLPVWPPTL 60
QY 61 VTTFSTGYOCFSRYPDHMKRHDFFKSAMPEGYVQOQRTIFPKDGNKTRAEVKFEEDTLV 120
DB 61 VTTFSTGYOCFSRYPDHMKRHDFFKSAMPEGYVQOQRTIFPKDGNKTRAEVKFEEDTLV 120
QY 121 NRTELKIDFKEDGNILGKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
DB 121 NRTELKIDFKEDGNILGKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 121 NRTELKIDFKEDGNILGKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
DB 121 NRTELKIDFKEDGNILGKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 181 YVOQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITGHMDELYK 238
DB 181 YVOQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITGHMDELYK 238

RESULT 11
US-09-852-000-1
; Sequence 1, Application US/09852000
; Patent No. US20020099170A1
; GENERAL INFORMATION:
; APPLICANT: Osumi, Takashi
; APPLICANT: Tsukamoto, Toshiro
; APPLICANT: Tsukamoto, No. US20020099170A1Iyo
; APPLICANT: Yamasaki, Masatoshi
; TITLE OF INVENTION: GREEN FLUORESCENT PROTEINS AND BLUE FLUORESCENT
; FILE REFERENCE: 046124-5005-US
; CURRENT APPLICATION NUMBER: US/09/852,000
; CURRENT FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: JP 026418/1998
; PRIOR FILING DATE: 1998-01-23
; PRIOR APPLICATION NUMBER: US 09/121,539
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 09/615,655
; PRIOR FILING DATE: 2000-07-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; OTHER INFORMATION: Green fluorescent protein
US-09-852-000-1

Query Match          97.8%; Score 1242; DB 10; Length 238;
Best Local Similarity 97.5%; Pred. No. 6,1e-111;
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGELFTAVVPIVLVDGVDVNGHKFVSVSGEGEDVYTKLTKFKICTTGKLPVWPPTL 60
```

```

DB 1 MSKGELFTGVVPIVLVDGVDVNGHKFVSVSGEGEDATYKTLTKFKICTTGKLPVWPPTL 60
QY 61 VTTFSTGYOCFSRYPDHMKRHDFFKSAMPEGYVQOQRTIFPKDGNKTRAEVKFEEDTLV 120
DB 61 VTTFSTGYOCFSRYPDHMKRHDFFKSAMPEGYVQOQRTIFPKDGNKTRAEVKFEEDTLV 120
QY 121 NRTELKIDFKEDGNILGKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
DB 121 NRTELKIDFKEDGNILGKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 121 NRTELKIDFKEDGNILGKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
DB 121 NRTELKIDFKEDGNILGKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 181 YVOQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITGHMDELYK 238
DB 181 YVOQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITGHMDELYK 238

RESULT 12
US-09-900-345A-60
; Sequence 60, Application US/09900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; FILE REFERENCE: 10338-505
; CURRENT APPLICATION NUMBER: US/09/900,345A
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU PB8078
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/00008
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 60
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Leu(CTA)5Gpp
US-09-900-345A-60

Query Match          97.6%; Score 1239; DB 9; Length 243;
Best Local Similarity 97.1%; Pred. No. 1,2e-110;
Matches 231; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGELFTAVVPIVLVDGVDVNGHKFVSVSGEGEDVYTKLTKFKICTTGKLPVWPPTL 60
DB 6 LSKGELFTGVVPIVLVDGVDVNGHKFVSVSGEGEDATYKTLTKFKICTTGKLPVWPPTL 65
QY 61 VTTFSTGYOCFSRYPDHMKRHDFFKSAMPEGYVQOQRTIFPKDGNKTRAEVKFEEDTLV 120
DB 61 VTTFSTGYOCFSRYPDHMKRHDFFKSAMPEGYVQOQRTIFPKDGNKTRAEVKFEEDTLV 125
QY 121 NRTELKIDFKEDGNILGKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
DB 121 NRTELKIDFKEDGNILGKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 185
QY 126 NRTELKIDFKEDGNILGKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 185
DB 181 YVOQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITGHMDELYK 238
DB 186 YVOQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITGHMDELYK 243

RESULT 13
US-09-900-345A-62
; Sequence 62, Application US/09900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; FILE REFERENCE: 10338-505
```

```

: CURRENT APPLICATION NUMBER: US/09/900.345A
: CURRENT FILING DATE: 2001-07-06
: PRIOR APPLICATION NUMBER: AU PP8078
: PRIOR FILING DATE: 1999-01-08
: PRIOR APPLICATION NUMBER: PCT/AU00/00008
: PRIOR FILING DATE: 2000-01-07
: NUMBER OF SEQ ID NOS: 185
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 62
: LENGTH: 243
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Leu(CTC)5GFP
US-09-900-345A-62
```

```

Query Match          97.6%; Score 1239; DB 9; Length 243;
Best Local Similarity 97.1%; Pred. No. 1.2e-110;
Matches 231; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```

QY 1 MSKGEELFTAVNPILVELDGDVNGHKFSVSGEGSDVYTKLTKFICTTGKLPVMPPTL 60
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 6 LSKGEELFTGVNPILVELDGDVNGHKFSVSGEGSDATYKTLTKFICTTGKLPVMPPTL 65
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 61 VTTFSTGYOCFSRYPDHMKRHDFFKSAPEGYVOQRTIFFKDGNGYKTRAEVKEGDTLV 120
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 66 VTTFSTGYOCFSRYPDHMKRHDFFKSAPEGYVOERTIFFKDGNGYKTRAEVKEGDTLV 125
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 121 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVFKIRHNIEDGSVOLAD 180
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 126 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVFKIRHNIEDGSVOLAD 185
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 181 YYOQNPILDGPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITHGMDLYK 238
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 186 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITHGMDLYK 243
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
```

```

RESULT 14
US-09-900-345A-64
: Sequence 64, Application US/09900345A
: Publication No. US20030031999A1
: GENERAL INFORMATION:
: APPLICANT: Frazer, Ian Hector
: TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
: FILE REFERENCE: 10338-5US
: CURRENT APPLICATION NUMBER: US/09/900.345A
: CURRENT FILING DATE: 2001-07-06
: PRIOR APPLICATION NUMBER: AU PP8078
: PRIOR FILING DATE: 1999-01-08
: PRIOR APPLICATION NUMBER: PCT/AU00/00008
: PRIOR FILING DATE: 2000-01-07
: NUMBER OF SEQ ID NOS: 185
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 64
: LENGTH: 243
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Leu(CTG)5GFP
US-09-900-345A-64
```

```

Query Match          97.6%; Score 1239; DB 9; Length 243;
Best Local Similarity 97.1%; Pred. No. 1.2e-110;
Matches 231; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```

QY 1 MSKGEELFTAVNPILVELDGDVNGHKFSVSGEGSDVYTKLTKFICTTGKLPVMPPTL 60
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 6 LSKGEELFTGVNPILVELDGDVNGHKFSVSGEGSDATYKTLTKFICTTGKLPVMPPTL 65
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 61 VTTFSTGYOCFSRYPDHMKRHDFFKSAPEGYVOQRTIFFKDGNGYKTRAEVKEGDTLV 120
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
```

```

DB 66 VTTFSTGYOCFSRYPDHMKRHDFFKSAPEGYVOERTIFFKDGNGYKTRAEVKEGDTLV 125
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 121 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVFKIRHNIEDGSVOLAD 180
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 126 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVFKIRHNIEDGSVOLAD 185
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 181 YYOQNPILDGPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITHGMDLYK 238
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 186 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITHGMDLYK 243
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
```

```

RESULT 15
US-09-900-345A-66
: Sequence 66, Application US/09900345A
: Publication No. US20030031999A1
: GENERAL INFORMATION:
: APPLICANT: Frazer, Ian Hector
: TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
: FILE REFERENCE: 10338-5US
: CURRENT APPLICATION NUMBER: US/09/900.345A
: CURRENT FILING DATE: 2001-07-06
: PRIOR APPLICATION NUMBER: AU PP8078
: PRIOR FILING DATE: 1999-01-08
: PRIOR APPLICATION NUMBER: PCT/AU00/00008
: PRIOR FILING DATE: 2000-01-07
: NUMBER OF SEQ ID NOS: 185
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 66
: LENGTH: 243
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Leu(CTT)5GFP
US-09-900-345A-66
```

```

Query Match          97.6%; Score 1239; DB 9; Length 243;
Best Local Similarity 97.1%; Pred. No. 1.2e-110;
Matches 231; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```

QY 1 MSKGEELFTAVNPILVELDGDVNGHKFSVSGEGSDVYTKLTKFICTTGKLPVMPPTL 60
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 6 LSKGEELFTGVNPILVELDGDVNGHKFSVSGEGSDATYKTLTKFICTTGKLPVMPPTL 65
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 61 VTTFSTGYOCFSRYPDHMKRHDFFKSAPEGYVOQRTIFFKDGNGYKTRAEVKEGDTLV 120
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 66 VTTFSTGYOCFSRYPDHMKRHDFFKSAPEGYVOERTIFFKDGNGYKTRAEVKEGDTLV 125
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 121 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVFKIRHNIEDGSVOLAD 180
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 126 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKVFKIRHNIEDGSVOLAD 185
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 181 YYOQNPILDGPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITHGMDLYK 238
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
DB 186 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITHGMDLYK 243
    :|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
```

Search completed: July 11, 2003, 18:08:22
Job time : 54 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 17:55:34 ; Search time 27 Seconds
(without alignments)
259.358 Million cell updates/sec

Title: US-10-071-976-2

Perfect score: 1270

Sequence: 1 MSKGEELFAVAVPIVLVDG.....VLLEFVTAAGITHGMDLYK 238

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_AA:*

1: /cgn2.6/ptodata/1/1aa/5A.COMB.pep:*

2: /cgn2.6/ptodata/1/1aa/5B.COMB.pep:*

3: /cgn2.6/ptodata/1/1aa/6A.COMB.pep:*

4: /cgn2.6/ptodata/1/1aa/6B.COMB.pep:*

5: /cgn2.6/ptodata/1/1aa/PT05.COMB.pep:*

6: /cgn2.6/ptodata/1/1aa/Backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1270	100.0	238	3	US-08-911-825-2
2	1270	100.0	238	3	US-08-974-737-2
3	1270	100.0	238	4	US-08-706-408-2
4	1270	100.0	238	4	US-09-465-142-2
5	1246	98.1	238	1	US-08-753-143-2
6	1246	98.1	238	2	US-08-679-865-2
7	1246	98.1	238	2	US-08-680-876-2
8	1246	98.1	238	2	US-08-792-553-2
9	1246	98.1	238	3	US-08-753-144-2
10	1246	98.1	238	4	US-09-094-359-2
11	1246	98.1	238	4	US-09-172-063-2
12	1246	98.1	238	4	US-09-263-975-2
13	1246	98.1	238	4	US-08-727-452-2
14	1246	98.1	238	4	US-08-418-785-1
15	1242	97.8	238	1	US-08-337-915A-2
16	1242	97.8	238	4	US-09-121-539-1
17	1242	97.8	238	5	PCF-US95-14692-2
18	1241	97.7	239	3	US-08-646-538-2
19	1241	97.7	239	4	US-09-503-222-2
20	1239	97.6	1070	4	US-09-091-042A-2
21	1239	97.6	1452	4	US-09-127-227-2
22	1233	97.1	238	3	US-08-893-327-16
23	1233	97.1	247	3	US-08-893-327-18
24	1233	97.1	247	3	US-08-893-327-20
25	1232	97.0	238	4	US-09-213-343-4
26	1230	96.9	238	4	US-08-588-201-2
27	1230	96.9	238	2	US-09-169-605-2

28	1230	96.9	238	3	US-08-893-327-2	Sequence 2, Appl1
29	1226	96.5	238	4	US-08-643-704A-49	Sequence 49, Appl1
30	1224	96.4	239	3	US-09-121-539-14	Sequence 14, Appl1
31	1221	96.1	238	1	US-08-452-295-1	Sequence 1, Appl1
32	1221	96.1	238	4	US-09-172-063-8	Sequence 8, Appl1
33	1221	96.1	238	4	US-09-316-919-9	Sequence 9, Appl1
34	1220	96.1	239	4	US-09-513-783A-52	Sequence 52, Appl1
35	1220	96.1	282	4	US-09-513-783A-14	Sequence 14, Appl1
36	1220	96.1	287	4	US-09-513-783A-8	Sequence 8, Appl1
37	1220	96.1	291	4	US-09-513-783A-10	Sequence 10, Appl1
38	1220	96.1	292	4	US-09-513-783A-16	Sequence 16, Appl1
39	1220	96.1	295	4	US-09-513-783A-20	Sequence 20, Appl1
40	1220	96.1	296	4	US-09-513-783A-12	Sequence 12, Appl1
41	1220	96.1	302	4	US-09-513-783A-18	Sequence 18, Appl1
42	1220	96.1	316	4	US-09-513-783A-26	Sequence 26, Appl1
43	1220	96.1	326	4	US-09-513-783A-24	Sequence 24, Appl1
44	1220	96.1	347	4	US-09-513-783A-38	Sequence 28, Appl1
45	1220	96.1	350	4	US-09-513-783A-30	Sequence 30, Appl1

ALIGNMENTS

RESULT 1
US-08-911-825-2
; Sequence 2, Application US/08911825
; Patent No. 6054321
; GENERAL INFORMATION:
; APPLICANT: Tsien et al., Roger Y. et al.
; TITLE OF INVENTION: LONG WAVELENGTH MUTANT FLUORESCENT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,825
; FILING DATE: 15-AUG-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hallie, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07257/056001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-911-825-2

Query Match 100.0%; Score 1270; DB 3; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.5e-130;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSKGEELFAVAVPIVLVDGNGHKEFSVSGEGEDVYTGKTLTFTGKLPVWPPL 60
|||||
DB 1 MSKGEELFAVAVPIVLVDGNGHKEFSVSGEGEDVYTGKTLTFTGKLPVWPPL 60
|||||
OY 61 VTFSGVGCFSRYPDHMKRHFESKAMEGYYOORTIFFKDDGNYKTRAEEKFEEDTLV 120
|||||

Db 61 VTTSYGVQCFSRPDMKRRDFKSAPEGYVOQRTIFFKDDGNKTRAEVKEGDTLV 120
QY 121 NRIELKIDFKEDGNILGHKLENYNSHNHYIMADKONGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRIELKIDFKEDGNILGHKLENYNSHNHYIMADKONGIKVFKIRHNIEDGSVOLAD 180
QY 181 YVOQNTPIIDGPVLLPDMHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITHGMDLYK 238
Db 181 YVOQNTPIIDGPVLLPDMHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITHGMDLYK 238

RESULT 2

US-08-974-737-2
Sequence 2, Application US/08974737
Patent No. 6077707
GENERAL INFORMATION:
APPLICANT: Tsien et al., Roger Y. et al.
TITLE OF INVENTION: LONG WAVELENGTH MUTANT FLUORESCENT
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,737
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/911,825
FILING DATE: 15-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07257/056001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-974-737-2

Query Match 100.0%; Score 1270; DB 3; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.5e-130;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVNPILVELDGDVNGHKFSVSGEGSDVTYGLTKLFICTTGKLPVMPPTL 60
Db 1 MSKGEELFTAVNPILVELDGDVNGHKFSVSGEGSDVTYGLTKLFICTTGKLPVMPPTL 60
QY 61 VTTSYGVQCFSRPDMKRRDFKSAPEGYVOQRTIFFKDDGNKTRAEVKEGDTLV 120
Db 61 VTTSYGVQCFSRPDMKRRDFKSAPEGYVOQRTIFFKDDGNKTRAEVKEGDTLV 120
QY 121 NRIELKIDFKEDGNILGHKLENYNSHNHYIMADKONGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRIELKIDFKEDGNILGHKLENYNSHNHYIMADKONGIKVFKIRHNIEDGSVOLAD 180
QY 181 YVOQNTPIIDGPVLLPDMHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITHGMDLYK 238
Db 181 YVOQNTPIIDGPVLLPDMHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITHGMDLYK 238

RESULT 3
US-08-706-408-2
Sequence 2, Application US/08706408
Patent No. 6124128

GENERAL INFORMATION:
APPLICANT: Tsien et al., Roger Y.
TITLE OF INVENTION: LONG WAVELENGTH MUTANT FLUORESCENT
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/706,408
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,050
FILING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07257/029001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-706-408-2

Query Match 100.0%; Score 1270; DB 3; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.5e-130;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVNPILVELDGDVNGHKFSVSGEGSDVTYGLTKLFICTTGKLPVMPPTL 60
Db 1 MSKGEELFTAVNPILVELDGDVNGHKFSVSGEGSDVTYGLTKLFICTTGKLPVMPPTL 60
QY 61 VTTSYGVQCFSRPDMKRRDFKSAPEGYVOQRTIFFKDDGNKTRAEVKEGDTLV 120
Db 61 VTTSYGVQCFSRPDMKRRDFKSAPEGYVOQRTIFFKDDGNKTRAEVKEGDTLV 120
QY 121 NRIELKIDFKEDGNILGHKLENYNSHNHYIMADKONGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRIELKIDFKEDGNILGHKLENYNSHNHYIMADKONGIKVFKIRHNIEDGSVOLAD 180
QY 181 YVOQNTPIIDGPVLLPDMHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITHGMDLYK 238
Db 181 YVOQNTPIIDGPVLLPDMHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITHGMDLYK 238

RESULT 4
US-09-465-142-2
Sequence 2, Application US/09465142
Patent No. 6403374
GENERAL INFORMATION:
APPLICANT: Tsien et al., Roger Y. et al.
TITLE OF INVENTION: LONG WAVELENGTH MUTANT FLUORESCENT

```

; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/465,142
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/974,737
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07257/056001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-465-142-2

```

```

Query Match      100.0%; Score 1270; DB 4; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.5e-130;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MSKGEELFTAVVPILVELDGVNGHKFSVSGEGEDVYTGKTLTKFICTTGKLPVWPPTL 60
DB 1 MSKGEELFTAVVPILVELDGVNGHKFSVSGEGEDVYTGKTLTKFICTTGKLPVWPPTL 60
QY 61 VTFESYGVQCFSRYPDHMKRHDFFKSAPEGVVQORTIFFKDDGNVKTAEVKEFGDITLV 120
DB 61 VTFESYGVQCFSRYPDHMKRHDFFKSAPEGVVQORTIFFKDDGNVKTAEVKEFGDITLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNYNSHNVYIMADKOKNGIKVNFKIRHNIEDGSVQLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNNYNSHNVYIMADKOKNGIKVNFKIRHNIEDGSVQLAD 180
QY 181 YVQONTPIIDGFPVLLPDNHYLSTQSALSKDPNEKRDHMYLLEFVTAAGITHGMDELTK 238
DB 181 YVQONTPIIDGFPVLLPDNHYLSTQSALSKDPNEKRDHMYLLEFVTAAGITHGMDELTK 238

```

```

RESULT 5
US-08-753-143-2
; Sequence 2, Application US/08753143A
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Heim, Roger
; TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
; FILE REFERENCE: 07257/032003
; CURRENT APPLICATION NUMBER: US/08/753,143A
; CURRENT FILING DATE: 1996-11-20
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria

```

```

US-08-753-143-2

```

```

Query Match      98.1%; Score 1246; DB 1; Length 238;
Best Local Similarity 97.9%; Pred. No. 6.1e-128;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 MSKGEELFTAVVPILVELDGVNGHKFSVSGEGEDVYTGKTLTKFICTTGKLPVWPPTL 60
DB 1 MSKGEELFTAVVPILVELDGVNGHKFSVSGEGEDVYTGKTLTKFICTTGKLPVWPPTL 60
QY 61 VTFESYGVQCFSRYPDHMKRHDFFKSAPEGVVQORTIFFKDDGNVKTAEVKEFGDITLV 120
DB 61 VTFESYGVQCFSRYPDHMKRHDFFKSAPEGVVQORTIFFKDDGNVKTAEVKEFGDITLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNYNSHNVYIMADKOKNGIKVNFKIRHNIEDGSVQLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNNYNSHNVYIMADKOKNGIKVNFKIRHNIEDGSVQLAD 180
QY 181 YVQONTPIIDGFPVLLPDNHYLSTQSALSKDPNEKRDHMYLLEFVTAAGITHGMDELTK 238
DB 181 YVQONTPIIDGFPVLLPDNHYLSTQSALSKDPNEKRDHMYLLEFVTAAGITHGMDELTK 238

```

```

RESULT 6
US-08-679-865-2
; Sequence 2, Application US/08679865
; Patent No. 5912137
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Cubitt, Andrew B.
; TITLE OF INVENTION: Assays for Protein Kinases Using
; TITLE OF INVENTION: Fluorescent Protein Substrates
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 16-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Storella, John S.
; REGISTRATION NUMBER: 32,944
; REFERENCE/DOCKET NUMBER: 023072-069000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-679-865-2

```

```

Query Match      98.1%; Score 1246; DB 2; Length 238;
Best Local Similarity 97.9%; Pred. No. 6.1e-128;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 MSKGEELFTAVVPILVELDGVNGHKFSVSGEGEDVYTGKTLTKFICTTGKLPVWPPTL 60
DB 1 MSKGEELFTAVVPILVELDGVNGHKFSVSGEGEDVYTGKTLTKFICTTGKLPVWPPTL 60
QY 61 VTFESYGVQCFSRYPDHMKRHDFFKSAPEGVVQORTIFFKDDGNVKTAEVKEFGDITLV 120

```

|||||
Db 61 VTTSYGVQCFSRYPDHMKRHDFFKSAPEGYVQERTFFEKDDGNYKTRAVERKEGDTLV 120
OY 121 NRIELKGIIDFKEDGNILGHKLEIYNNNSHNHYIMADKONGIKVFKIRHNIEDSSVOLAD 180
Db 121 NRIELKGIIDFKEDGNILGHKLEIYNNNSHNHYIMADKONGIKVFKIRHNIEDSSVOLAD 180
OY 181 YVOQNTPIIDGPVLLPDPNHVLTQSALSADPNEKRDMHVLLEFYTAAGITGHMDELKX 238
Db 181 HYQONTPIIDGPVLLPDPNHVLTQSALSADPNEKRDMHVLLEFYTAAGITGHMDELKX 238

RESULT 7

US-08-680-876-2
; Sequence 2, Application US/08680876
; Patent No. 5925558
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Cubitt, Andrew B.
; TITLE OF INVENTION: Assays for Protein Kinases Using
; TITLE OF INVENTION: Fluorescent Protein Substrates
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680, 876
; FILING DATE: 16-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Storella, John S.
; REGISTRATION NUMBER: 32,944
; REFERENCE/DOCKET NUMBER: 02307Z-069200
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-680-876-2

Query Match 98.1%; Score 1246; DB 2; Length 238;
Best Local Similarity 97.9%; Pred. No. 6.1e-128;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSKGEELFTAVVPLVPLVELDGVNGHKFSVSGEGEDVYTGKLTAKTCTTGKLPVPVPTL 60
Db 1 MSKGEELFTGVPLVPLVELDGVNGHKFSVSGEGEDVYTGKLTAKTCTTGKLPVPVPTL 60
OY 61 VTTSYGVQCFSRYPDHMKRHDFFKSAPEGYVQERTFFEKDDGNYKTRAVERKEGDTLV 120
Db 61 VTTSYGVQCFSRYPDHMKRHDFFKSAPEGYVQERTFFEKDDGNYKTRAVERKEGDTLV 120
OY 121 NRIELKGIIDFKEDGNILGHKLEIYNNNSHNHYIMADKONGIKVFKIRHNIEDSSVOLAD 180
Db 121 NRIELKGIIDFKEDGNILGHKLEIYNNNSHNHYIMADKONGIKVFKIRHNIEDSSVOLAD 180
OY 181 YVOQNTPIIDGPVLLPDPNHVLTQSALSADPNEKRDMHVLLEFYTAAGITGHMDELKX 238
Db 181 HYQONTPIIDGPVLLPDPNHVLTQSALSADPNEKRDMHVLLEFYTAAGITGHMDELKX 238

RESULT 8

US-08-792-553-2
; Sequence 2, Application US/08792553
; Patent No. 5981200
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Helm, Roger
; TITLE OF INVENTION: Tandem Fluorescent Protein Constructs
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & RICHARDSON P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/792,553
; FILING DATE: 31-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lisa A. Halle, Ph.D.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07257/041001/DC 96-160-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-678-5070
; TELEFAX: 619-678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-792-553-2

Query Match 98.1%; Score 1246; DB 2; Length 238;
Best Local Similarity 97.9%; Pred. No. 6.1e-128;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSKGEELFTAVVPLVPLVELDGVNGHKFSVSGEGEDVYTGKLTAKTCTTGKLPVPVPTL 60
Db 1 MSKGEELFTGVPLVPLVELDGVNGHKFSVSGEGEDVYTGKLTAKTCTTGKLPVPVPTL 60
OY 61 VTTSYGVQCFSRYPDHMKRHDFFKSAPEGYVQERTFFEKDDGNYKTRAVERKEGDTLV 120
Db 61 VTTSYGVQCFSRYPDHMKRHDFFKSAPEGYVQERTFFEKDDGNYKTRAVERKEGDTLV 120
OY 121 NRIELKGIIDFKEDGNILGHKLEIYNNNSHNHYIMADKONGIKVFKIRHNIEDSSVOLAD 180
Db 121 NRIELKGIIDFKEDGNILGHKLEIYNNNSHNHYIMADKONGIKVFKIRHNIEDSSVOLAD 180
OY 181 YVOQNTPIIDGPVLLPDPNHVLTQSALSADPNEKRDMHVLLEFYTAAGITGHMDELKX 238
Db 181 HYQONTPIIDGPVLLPDPNHVLTQSALSADPNEKRDMHVLLEFYTAAGITGHMDELKX 238

RESULT 9

US-08-753-144-2
; Sequence 2, Application US/08753144
; Patent No. 6066476
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Helm, Roger
; TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.

```

; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/753,144
; FILING DATE: 20-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/727,452
; FILING DATE: 10-OCT-1996
; APPLICATION NUMBER: US95/14692
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: 08/337,915
; FILING DATE: 10-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07257/032002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-753-144-2

Query Match          98.1%; Score 1246; DB 3; Length 238;
Best Local Similarity 97.9%; Pred. No. 6,1e-128;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPIVLVDGNGHKFSVSGEGDVTYGGKLTFTCTTGKLPVPMPTL 60
DB 1 MSKGEELFTGVVPIVLVDGNGHKFSVSGEGDVTYGGKLTFTCTTGKLPVPMPTL 60
QY 61 VTTFSYGVQCFSRYPDHKKRHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120
DB 61 VTTFSYGVQCFSRYPDHKKRHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120
QY 121 NRIELKGIDFEKEDNIIHGKLELYNNSHNHYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFEKEDNIIHGKLELYNNSHNHYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
QY 181 YVQONTPLDGPVLLPDNHYLSTQSALS KDPNEKRDMHWLLEFVTAAGITGMDLYK 238
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHWLLEFVTAAGITGMDLYK 238

RESULT 10
US-09-094-359-2
; Sequence 2, Application US/09094359
; Patent No. 6140132
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/067001
; CURRENT APPLICATION NUMBER: US/09/094,359
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2

```

```

; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
; US-09-094-359-2

Query Match          98.1%; Score 1246; DB 4; Length 238;
Best Local Similarity 97.9%; Pred. No. 6,1e-128;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPIVLVDGNGHKFSVSGEGDVTYGGKLTFTCTTGKLPVPMPTL 60
DB 1 MSKGEELFTGVVPIVLVDGNGHKFSVSGEGDVTYGGKLTFTCTTGKLPVPMPTL 60
QY 61 VTTFSYGVQCFSRYPDHKKRHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120
DB 61 VTTFSYGVQCFSRYPDHKKRHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120
QY 121 NRIELKGIDFEKEDNIIHGKLELYNNSHNHYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFEKEDNIIHGKLELYNNSHNHYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
QY 181 YVQONTPLDGPVLLPDNHYLSTQSALS KDPNEKRDMHWLLEFVTAAGITGMDLYK 238
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHWLLEFVTAAGITGMDLYK 238

RESULT 11
US-09-172-063-2
; Sequence 2, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; EARLIER FILING DATE: 1998-10-13
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
; US-09-172-063-2

Query Match          98.1%; Score 1246; DB 4; Length 238;
Best Local Similarity 97.9%; Pred. No. 6,1e-128;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPIVLVDGNGHKFSVSGEGDVTYGGKLTFTCTTGKLPVPMPTL 60
DB 1 MSKGEELFTGVVPIVLVDGNGHKFSVSGEGDVTYGGKLTFTCTTGKLPVPMPTL 60
QY 61 VTTFSYGVQCFSRYPDHKKRHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120
DB 61 VTTFSYGVQCFSRYPDHKKRHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120
QY 121 NRIELKGIDFEKEDNIIHGKLELYNNSHNHYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFEKEDNIIHGKLELYNNSHNHYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
QY 181 YVQONTPLDGPVLLPDNHYLSTQSALS KDPNEKRDMHWLLEFVTAAGITGMDLYK 238
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHWLLEFVTAAGITGMDLYK 238

RESULT 12

```



```
QY      61 VTTFSSYGVQCFSRYPDHMKRHDFFKSAMPEGYVOARTIFPKDDGNKYKTRAEVKFEQDTLV 120
        |||
Db       61 VTTFSSYGVQCFSRYPDHMKRHDFFKSAMPEGYVOERTIFFKDDGNKYKTRAEVKFEQDTLV 120
        |||
QY      121 NRIELKGIIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
        |||
Db       121 NRIELKGIIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
        |||
QY      181 YVOQNTPIIDGPPVLLPDNHYLSTOSALSKDPEKRDHMYLLEFVTAAGITHGMDLYK 238
        :|||
Db       181 HYQONTPIIDGPPVLLPDNHYLSTOSALSKDPEKRDHMYLLEFVTAAGITHGMDLYK 238
        :|||
```

RESULT 15

```
US-08-337-915A-2
: Sequence 2, Application US/08337915A
: Patent No. 5625048
: GENERAL INFORMATION:
: APPLICANT: Tsien, Roger Y.
: APPLICANT: Heim, Roger
: TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Robbins, Berliner & Carson
: STREET: 201 No. 5625048th Figueroa Street, Suite 500
: CITY: Los Angeles
: STATE: California
: COUNTRY: USA
: ZIP: 90012
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/337,915A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Spitals, John P.
: REGISTRATION NUMBER: 29,215
: REFERENCE/DOCKET NUMBER: 1279-178
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 977-1001
: TELEFAX: (213) 977-1003
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 238 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-337-915A-2
```

```
Query Match          97.8%; Score 1242; DB 1; Length 238;
Best Local Similarity 97.5%; Pred. No. 1.7e-127;
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 MSKGEELFTAVVPLVLVELDGVNGHKFSVSGEGSDVYTGKLTLEKICTTGKLPVPMPTL 60
        |||
Db       1 MSKGEELFTGVVPLVLVELDGVNGHKFSVSGEGSDATYTGKLTLEKICTTGKLPVPMPTL 60
        |||
QY      61 VTTFSSYGVQCFSRYPDHMKRHDFFKSAMPEGYVOARTIFFKDDGNKYKTRAEVKFEQDTLV 120
        |||
Db       61 VTTFSSYGVQCFSRYPDHMKRHDFFKSAMPEGYVOERTIFFKDDGNKYKTRAEVKFEQDTLV 120
        |||
QY      121 NRIELKGIIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
        |||
Db       121 NRIELKGIIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
        |||
QY      181 YVOQNTPIIDGPPVLLPDNHYLSTOSALSKDPEKRDHMYLLEFVTAAGITHGMDLYK 238
        :|||
Db       181 HYQONTPIIDGPPVLLPDNHYLSTOSALSKDPEKRDHMYLLEFVTAAGITHGMDLYK 238
        :|||
```

Search completed: July 11, 2003, 18:00:06
Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 17:51:43 ; Search time 81 Seconds

(Without alignments)
605.422 Million cell updates/sec

Title: US-10-071-976-2

Perfect score: 1270
Sequence: 1 MSKGEELFTAVVPLIVELDG.....VLEFVVAAGITGHMDELYK 238

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1242	97.8	238	5	Q27903
2	1235	97.2	238	5	Q93125
3	1200	94.5	238	5	Q17105
4	1185	93.3	238	5	Q17106
5	1182	93.1	758	4	Q96J01
6	1079	85.0	238	5	Q8WP95
7	1073	84.5	238	5	Q8WTC6
8	1065	83.9	238	5	Q8WTC4
9	1063	83.7	238	5	Q8WTC0
10	1062	83.6	238	5	Q8WTC9
11	1060	83.6	238	5	Q8WTC8
12	1058	83.3	238	5	Q8WTC5
13	1058	83.3	238	5	Q8WTC5
14	255.5	20.1	225	5	Q95U47
15	245	19.3	225	5	Q963F5
16	241.5	19.0	266	5	Q906Y3

17	239.5	18.9	225	5	Q8T5F1	Q8T5F1 montastraea
18	239.5	18.9	236	5	Q8T5R0	Q8T5R0 dendronephth
19	212.5	16.7	235	5	Q8T5R0	Q8T5R0 scolymia cu
20	212.5	16.7	238	5	Q9BLT9	Q9BLT9 renilla mu
21	209.5	16.5	225	5	Q9U6Y8	Q9U6Y8 discosoma s
22	209	16.5	229	5	Q9U6Y6	Q9U6Y6 anemonia ma
23	206.5	16.3	234	5	Q8T5F2	Q8T5F2 montastraea
24	205	16.1	232	5	Q9U6Y7	Q9U6Y7 discosoma s
25	204	16.1	227	5	Q962P9	Q962P9 montastraea
26	203	16.0	221	5	Q95P04	Q95P04 gonioporta t
27	200.5	15.8	232	5	Q9GP15	Q9GP15 anemonia su
28	200	15.7	227	5	Q95VW0	Q95VW0 montastraea
29	199.5	15.7	225	5	Q8T6P9	Q8T6P9 heteractis
30	196.5	15.5	232	5	Q9GZT8	Q9GZT8 anemonia su
31	196	15.4	230	5	Q9GZT7	Q9GZT7 discosoma s
32	191.5	15.1	233	5	Q963I9	Q963I9 renilla ren
33	190.5	15.0	238	5	Q9BLT0	Q9BLT0 pillosarcus
34	189.5	14.9	228	5	Q9GP16	Q9GP16 anemonia su
35	183.5	14.4	227	5	Q95W85	Q95W85 heteractis
36	183.5	14.4	231	5	Q9U6Y5	Q9U6Y5 zoanthus sp
37	176.5	13.9	229	5	Q8T5E7	Q8T5E7 condylactis
38	175.5	13.8	227	5	Q95W86	Q95W86 condylactis
39	173	13.6	231	5	Q9U6Y4	Q9U6Y4 zoanthus sp
40	169.5	13.3	227	5	Q95W11	Q95W11 condylactis
41	168.5	13.3	231	5	Q8T4D4	Q8T4D4 zoanthus sp
42	168	13.2	234	5	Q8T5F3	Q8T5F3 scolymia cu
43	165	13.0	231	5	Q8T5F8	Q8T5F8 ricordea fl
44	159.5	12.6	231	5	Q8T6T8	Q8T6T8 discosoma s
45	151	11.9	231	5	Q8T5E9	Q8T5E9 ricordea fl

ALIGNMENTS

RESULT 1

ID Q27903 PRELIMINARY: PRT, 238 AA.
AC Q27903:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS unidentified.
OC unclassified.
OX NCBI_TaxID=32644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97299832; PubMed=9154981;
RA Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RT "Enhanced expression in tobacco of the gene encoding green fluorescent
protein by modification of its codon usage."
RL Plant Mol. Biol. 33:989-999(1997).
DR EMBL; X96418; CAA65278.1; -
DR HSSP; P42212; IGFL.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP.1.
DR PRINTS; PR01229; GFPDORESCENT.
DR PRODom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26884 MW; CA932D47262AF2D3 CRC64;

Query Match 97.8%; Score 1242; DB 5; Length 238;
Best Local Similarity 97.5%; Pred. No. 7.7e-94;
Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY	1	MSKGEELFTAVVPLIVELDGVNGHKFSVSGEGSDVYTGKLTFFICTGKLPVMPPTL 60
DB	1	MSKGEELFTAVVPLIVELDGVNGHKFSVSGEGSDVYTGKLTFFICTGKLPVMPPTL 60
QY	61	VTTTSYGVQCFSRPDDHKKRDFRSAMPBGVVOORTFFFDGDKYKRAVKEGDTLV 120
DB	61	VTTTSYGVQCFSRPDDHKKRDFRSAMPBGVVOORTFFFDGDKYKRAVKEGDTLV 120

QY 121 NRLEKIDKEKGNIIIGHKLEYNNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 DB 121 NRLEKIDKEKGNIIIGHKLEYNNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 QY 181 YVOONTPIIDGVPVLLPDNHYLSTOSALSKDPNEKRDMHVLLEFVTAAGITHGMDLYK 238
 DB 181 HYQONTPIIDGVPVLLPDNHYLSTOSALSKDPNEKRDMHVLLEFVTAAGITHGMDLYK 238

RESULT 2

ID 093125 PRELIMINARY; PRT; 238 AA.
 AC 093125;
 DT 01-FEB-1997 (Tremblrel. 02, Created)
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Green fluorescent protein mutant 3.
 GN GFP.
 OS Aequorea victoria (Jellyfish).
 OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;
 OC Aequoreidae; Aequorea.
 OX NCBI_TaxID=6100;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96305137; PubMed=8707053;
 RA Cormack B.P., Valdivia R.H., Falkow S.;
 RT "FACS-Optimized mutants of the green fluorescent protein (GFP).";
 RL Gene 173:33-38(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cormack B.P., Egerton M., Gow N.A.R., Falkow S.;
 RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene
 expression in Candida albicans";
 RL Microbiology 0:0-0(1996).
 DR EMBL: U73901; AAB18957.1; -.
 DR HSSP: P42212; 1BFP.
 DR InterPro: IPR000786; Green_fl_protein.
 DR Pfam: PF01353; GFP; 1.
 DR PRINTS: PR01229; GFPLORESCENT.
 DR PRODOM: PD013756; Green_fl_protein; 1.
 SQ SEQUENCE 238 AA; 26840 MW; A28622809A3DEA60 CRC64;

Query Match 97.2%; Score 1235; DB 5; Length 238;
 Best Local Similarity 96.6%; Pred. No. 2.9e-93;
 Matches 230; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPIVLVDGVDVNGHKFSVSGEGSDVTYGLTLKFICTTGKLPVMPPTL 60
 DB 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGSDVTYGLTLKFICTTGKLPVMPPTL 60
 QY 61 VTFESYGVQCFSRYPDHMKRHDFEFKSAMPEGYVQARTIFFKDDGNKTRAEVKFGDPTLV 120
 DB 61 VTFESYGVQCFARYPDHMKRHDFEFKSAMPEGYVQERTIFFKDDGNKTRAEVKFGDPTLV 120
 QY 121 NRLEKIDKEKGNIIIGHKLEYNNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 DB 121 NRLEKIDKEKGNIIIGHKLEYNNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 QY 181 YVOONTPIIDGVPVLLPDNHYLSTOSALSKDPNEKRDMHVLLEFVTAAGITHGMDLYK 238
 DB 181 HYQONTPIIDGVPVLLPDNHYLSTOSALSKDPNEKRDMHVLLEFVTAAGITHGMDLYK 238

RESULT 3

ID 017105 PRELIMINARY; PRT; 238 AA.
 AC 017105;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Green fluorescent protein (Fragment).
 GN GFP.

OS Aequorea victoria (Jellyfish).
 OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;
 OC Aequoreidae; Aequorea.
 OX NCBI_TaxID=6100;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Watkins J.N., Campbell A.K.;
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X83959; CAA58789.1; -.
 DR HSSP: P42212; 1GFL.
 DR InterPro: IPR000786; Green_fl_protein.
 DR Pfam: PF01353; GFP; 1.
 DR PRINTS: PR01229; GFPLORESCENT.
 DR PRODOM: PD013756; Green_fl_protein; 1.
 FT NON_TER 238
 SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 94.5%; Score 1200; DB 5; Length 238;
 Best Local Similarity 93.3%; Pred. No. 2.1e-90;
 Matches 222; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPIVLVDGVDVNGHKFSVSGEGSDVTYGLTLKFICTTGKLPVMPPTL 60
 DB 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGSDVTYGLTLKFICTTGKLPVMPPTL 60
 QY 61 VTFESYGVQCFSRYPDHMKRHDFEFKSAMPEGYVQARTIFFKDDGNKTRAEVKFGDPTLV 120
 DB 61 VTFESYGVQCFSRYPDHMKRHDFEFKSAMPEGYVQERTIFFKDDGNKTRAEVKFGDPTLV 120
 QY 121 NRLEKIDKEKGNIIIGHKLEYNNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 DB 121 NRLEKIDKEKGNIIIGHKLEYNNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 QY 181 YVOONTPIIDGVPVLLPDNHYLSTOSALSKDPNEKRDMHVLLEFVTAAGITHGMDLYK 238
 DB 181 HYQONTPIIDGVPVLLPDNHYLSTOSALSKDPNEKRDMHVLLEFVTAAGITHGMDLYK 238

RESULT 4

ID 017106 PRELIMINARY; PRT; 238 AA.
 AC 017106;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Green fluorescent protein (Fragment).
 GN GFP.
 OS Aequorea victoria (Jellyfish).
 OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;
 OC Aequoreidae; Aequorea.
 OX NCBI_TaxID=6100;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Watkins J.N., Campbell A.K.;
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X83960; CAA58790.1; -.
 DR HSSP: P42212; 1BFP.
 DR InterPro: IPR000786; Green_fl_protein.
 DR Pfam: PF01353; GFP; 1.
 DR PRINTS: PR01229; GFPLORESCENT.
 DR PRODOM: PD013756; Green_fl_protein; 1.
 FT NON_TER 238
 SQ SEQUENCE 238 AA; 26867 MW; BD4648262DBEABD4 CRC64;

Query Match 93.3%; Score 1185; DB 5; Length 238;
 Best Local Similarity 92.4%; Pred. No. 3.5e-89;
 Matches 220; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPIVLVDGVDVNGHKFSVSGEGSDVTYGLTLKFICTTGKLPVMPPTL 60
 DB 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGSDVTYGLTLKFICTTGKLPVMPPTL 60
 QY 61 VTFESYGVQCFSRYPDHMKRHDFEFKSAMPEGYVQARTIFFKDDGNKTRAEVKFGDPTLV 120

```

Db      61 VTFESYGVOCFSRYPDHKHNDHFKLSAMEGEGVVOERTIFYKXDDGNYKTRAEVKFGSDTLV 120
        |||
QY      121 NRTELKIDFKEDGNILGHKLEYNNSHNHYIMAKONGKIVNKRIRHNIDEGSVQIAD 180
        |||
Db      121 NRTELKIDFKEDGNILGHKLEYNNSHNHYIMSGKPNKIRNIRHNIDEGSVQIAD 180
        |||
QY      181 YVOQNTPIIDGFPVLLPDNHYISTOSALSQDPNEKRDHMYLLEFVTAAGITGMDLXK 238
        |||
Db      181 HYQONTPIIGDGFVLLPDNHYISTOSALSQDPNKRDKHMYLLEFVTAAGITGMDLXK 238
        |||

```

RESULT 5

```

096J01  PRELIMINARY; PRT; 758 AA.
ID AC 096J01
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Ralchu404X.
GN RAICHU404X.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RC TISSUE=SPLEEN.
RX MEDLINE=21322811; PubMed=11429608;
RA Mochizuki N., Yamashita S., Kurokawa K., Ohba Y., Nagai T.,
RA Miyawaki A., Matsuda M.;
RT "Spacio-temporal Images of Growth Factor-Induced Activation of Ras and
RT Rap1.";
RL Nature 411:1065-1068(2001).
DR EMBL; AB051846; BAB61868.1;
DR InterPro; IPR000786; Green_fl_protein.
DR InterPro; IPR001230; Prenyl_site.
DR InterPro; IPR001806; Ras_transfmrng.
DR InterPro; IPR003116; RBD.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF01353; GFP; 2.
DR Pfam; PF00071; ras; 1.
DR Pfam; PF02196; RBD; 1.
DR Prodom; PD013756; Green_fl_protein; 2.
DR TIGRFAMS; TIGR00231; small_GTP; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW GTP-binding.
SQ
SEQUENCE 758 AA; 85015 MW; 8612408F607CFD49 CRC64;

```

Query Match 93.1%; Score 1182; DB 4; Length 758;
 Best Local Similarity 94.1%; Pred. No. 2,7e-88;
 Matches 222; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

```

QY      1 MSKGEELFAVVPVILVELDGVNGHGFVSFSGEGEDVYTGKTLFICTGKLPVWPPTL 60
        |||
Db      499 VSKGEELFVGVPVILVELDGVNGHGFVSFSGEGEDATVGLKTLFICTGKLPVWPPTL 558
        |||
QY      61 VTFESYGVOCFSRYPDHKHNDHFKLSAMEGEGVVOERTIFYKXDDGNYKTRAEVKFGSDTLV 120
        |||
Db      559 VTLTWGVOCFSRYPDHKHNDHFKLSAMEGEGVVOERTIFYKXDDGNYKTRAEVKFGSDTLV 618
        |||
QY      121 NRTELKIDFKEDGNILGHKLEYNNSHNHYIMADKONGKIVNKRIRHNIDEGSVQIAD 180
        |||
Db      619 NRTELKIDFKEDGNILGHKLEYNNSHNHYITAAKONGKIRNIRHNIDEGSVQIAD 678
        |||
QY      181 YVOQNTPIIDGFPVLLPDNHYISTOSALSQDPNEKRDHMYLLEFVTAAGITGMDLXK 236
        |||
Db      679 HYQONTPIIGDGFVLLPDNHYISTOSALSQDPNEKRDHMYLLEFVTAAGITGMDLXK 734
        |||

```

RESULT 6

08WP95 PRELIMINARY; PRT; 238 AA.

```

AC      Q8WP95;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFPX.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_Taxid=147615;
RN [1]
RC STRAIN=GFPX, AND GFPDAXM;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qin Y.X., Pang S.O.,
RA Li S.J., Xia N.S.;
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY013824; AAK02052.1;
DR EMBL; AY013821; AAK02059.1;
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR Prodom; PD013756; Green_fl_protein; 1.
SQ
SEQUENCE 238 AA; 27049 MW; 8185D0E5E529012B CRC64;

```

Query Match 85.0%; Score 1079; DB 5; Length 238;
 Best Local Similarity 81.5%; Pred. No. 1.6e-80;
 Matches 194; Conservative 21; Mismatches 23; Indels 0; Gaps 0;

```

QY      1 MSKGEELFAVVPVILVELDGVNGHGFVSFSGEGEDVYTGKTLFICTGKLPVWPPTL 60
        |||
Db      1 MSKGEELFGLVPLVILVELDGVNGHGFVSFSGEGEDADVGKLEIFICTGKLPVWPPTL 60
        |||
QY      61 VTFESYGVOCFSRYPDHKHNDHFKLSAMEGEGVVOERTIFYKXDDGNYKTRAEVKFGSDTLV 120
        |||
Db      61 VTFESYGVOCFSRYPDHKHNDHFKLSAMEGEGVVOERTIFYKXDDGNYKTRAEVKFGSDTLV 120
        |||
QY      121 NRTELKIDFKEDGNILGHKLEYNNSHNHYIMADKONGKIVNKRIRHNIDEGSVQIAD 180
        |||
Db      121 NRTELKIDFKEDGNILGHKLEYNNSHNHYITMPDKANGKIVNKRIRHNIDEGSVQIAD 180
        |||
QY      181 YVOQNTPIIDGFPVLLPDNHYISTOSALSQDPNEKRDHMYLLEFVTAAGITGMDLXK 238
        |||
Db      181 HYQNTVPLDGFVLLPDNHYISTOSALSQDPNEKRDHMYLLEFVTAAGITGMDLXK 238
        |||

```

RESULT 7

```

ID AC 08WTC6
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_Taxid=147615;
RN [1]
RC STRAIN=GFPX19UV;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.O., Chen M.,
RA Li S.J., Xia N.S.;
RL "Colorful mutants of green fluorescent protein from Aequorea
RL macrodactyla.";
DR EMBL; AF435431; AAL33916.1;
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR Prodom; PD013756; Green_fl_protein; 1.
SQ
SEQUENCE 238 AA; 27015 MW; 6B88FD5E88926903 CRC64;

```

Query Match 84.5%; Score 1073; DB 5; Length 238;
 Best Local Similarity 81.1%; Pred. No. 5e-80;


```

QY 1 MSKGEELFTAVVPIVLVDGVDNGHKSFSVSGEGDVTYGGKLTLLKFICTTGKLPVWPPTL 60
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MSKGEELFTGIIVPLIELDGDVHGKFSVRGEGDADYGLKEIKFICTTGKLPVWPPTL 60
QY 61 VTTFSYGVQCSRRPDHKKRDFKRSAMPEGYVOORTFFDDGNGYKTRAAYKEEGDTLV 120
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 61 VTTGTYGIQCFARYPHEHKMDFKRSAMPEGYIOERTFFDDGKYKTRGVEKFGEDTLV 120
QY 121 NRIELKGIDFKEEDNIIIGHKLEYNYSNHNVIYIMADKONGIKVNFKIRHNIEDSGYQIAD 180
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 NRIELKGIDFKEEDNIIIGHKLEYNYSNHNVIYIMADKONGIKVNFKIRHNIEGGGYQIAD 180
QY 181 YYQONTPLDGPVLLPDNHYLSTOSALSKDPEKRDHNVLLFEVTAAGITGMDLYK 238
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 181 HYQNVNPLGDSGPVLLPINHYLSFQTAISKDRNETRDHNVFLFEFSAGCHTGMDELK 238

```

RESULT 11

```

Q8WTC8 PRELIMINARY; PRT; 238 AA.
ID 08WTC8
AC 08WTC8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
CX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPM163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
   Li S.J., Xia N.S.;
   "Colorful mutants of green fluorescent protein from Aequorea
   macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF435429; AAL33914.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

```

Query Match 83.6%; Score 1062; DB 5; Length 238;
 Best Local Similarity 80.3%; Pred. No. 4e-79;
 Matches 191; Conservative 21; Mismatches 26; Indels 0; Gaps 0;

```

QY 1 MSKGEELFTAVVPIVLVDGVDNGHKSFSVSGEGDVTYGGKLTLLKFICTTGKLPVWPPTL 60
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MSKGEELFTGIIVPLIELDGDVHGKFSVRGEGDADYGLKEIKFICTTGKLPVWPPTL 60
QY 61 VTTFSYGVQCSRRPDHKKRDFKRSAMPEGYVOORTFFDDGNGYKTRAAYKEEGDTLV 120
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 61 VTTGTYGIQCFARYPHEHKMDFKRSAMPEGYIOERTFFDDGKYKTRGVEKFGEDTLV 120
QY 121 NRIELKGIDFKEEDNIIIGHKLEYNYSNHNVIYIMADKONGIKVNFKIRHNIEDSGYQIAD 180
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 NRIELKGIDFKEEDNIIIGHKLEYNYSNHNVIYIMADKONGIKVNFKIRHNIEGGGYQIAD 180
QY 181 YYQONTPLDGPVLLPDNHYLSTOSALSKDPEKRDHNVLLFEVTAAGITGMDLYK 238
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 181 HYQNVNPLGDSGPVLLPINHYLSYQTAISKDRNETRDHNVFLFEFSAGCHTGMDELK 238

```

RESULT 12

```

Q8WTC7 PRELIMINARY; PRT; 238 AA.
ID 08WTC7
AC 08WTC7;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Green fluorescent protein.

```

```

GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
CX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPM191U;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
   Li S.J., Xia N.S.;
   "Colorful mutants of green fluorescent protein from Aequorea
   macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF435430; AAL33915.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27002 MW; BD5BA2982264C018 CRC64;

```

Query Match 83.5%; Score 1060; DB 5; Length 238;
 Best Local Similarity 80.3%; Pred. No. 5.8e-79;
 Matches 191; Conservative 21; Mismatches 26; Indels 0; Gaps 0;

```

QY 1 MSKGEELFTAVVPIVLVDGVDNGHKSFSVSGEGDVTYGGKLTLLKFICTTGKLPVWPPTL 60
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MSKGEELFTGIIVPLIELDGDVHGKFSVRGEGDADYGLKEIKFICTTGKLPVWPPTL 60
QY 61 VTTFSYGVQCSRRPDHKKRDFKRSAMPEGYVOORTFFDDGNGYKTRAAYKEEGDTLV 120
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 61 VTTGTYGIQCFARYPHEHKMDFKRSAMPEGYIOERTFFDDGKYKTRGVEKFGEDTLV 120
QY 121 NRIELKGIDFKEEDNIIIGHKLEYNYSNHNVIYIMADKONGIKVNFKIRHNIEDSGYQIAD 180
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 NRIELKGIDFKEEDNIIIGHKLEYNYSNHNVIYIMADKONGIKVNFKIRHNIEGGGYQIAD 180
QY 181 YYQONTPLDGPVLLPDNHYLSTOSALSKDPEKRDHNVLLFEVTAAGITGMDLYK 238
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 181 HYQNVNPLGDSGPVLLPINHYLSQTAISKDRNETRDHNVFLFEFSAGCHTGMDELK 238

```

RESULT 13

```

Q8WTC5 PRELIMINARY; PRT; 238 AA.
ID 08WTC5
AC 08WTC5;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Orange fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
CX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OPFXM;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
   Li S.J., Xia N.S.;
   "Colorful mutants of green fluorescent protein from Aequorea
   macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF435432; AAL33917.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 75521EA5534E573A CRC64;

```

Query Match 83.3%; Score 1058; DB 5; Length 238;
 Best Local Similarity 81.1%; Pred. No. 8.5e-79;
 Matches 193; Conservative 18; Mismatches 27; Indels 0; Gaps 0;

```

QY 1 MSKGEELFTAVVPIVLVDGVDNGHKSFSVSGEGDVTYGGKLTLLKFICTTGKLPVWPPTL 60
   |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

Db 1 MSKGEELFTGVPLVLELDGVDHGHKFSVSGEGSDADYKLEIKFICTTGKLPVWPPTL 60
 QY 61 VTFESYGVOCFSRYPDHMKRHFESKAMPBGVYOQRTIFFKDDGNYKTRAEVKEGGDTLV 120
 Db 61 VTLTGIGIICFARPEHMKNDFFSKAMPBGYIOERITFFQDDGKYTRREVKEGGDTLV 120
 QY 121 NRLELGIKIDFKEDGNILGHKLEYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 Db 121 NRLELGMDFEKEDGNILGHKLEYNSHNYIMDPKANNGLKVFKIRHNIEGGVQVLAD 180
 QY 181 YGOONPILDGVPVLPDNNHLSQSAISKDPNEKRDHMLLEFVTAAGITHGMDLTK 238
 Db 181 HYGTNPVLDGDPVLPINHYLSQTSALSKDNETRDMHMLLEFFSACGHTHGMDELTK 238

RESULT 14

095UA7 PRELIMINARY; PRT; 225 AA.

AC 095UA7; PRELIMINARY; PRT; 225 AA.
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Cyan fluorescent protein (Fragment).
 OS Montastraea cavernosa (great star coral).
 OC Eukaryota; Metazoa; Chordata; Anthozoa; Zoantharia; Scleractinia;
 OC Faviina; Faviidae; Montastraea.
 OX NCBI_TaxID=63558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Falkowski P.G., Sun Y.;
 RT "Montastraea cavernosa fluorescent protein.";
 RL Submitted (Sep-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY056460; AAL17905.1;
 DR InterPro; IPR000786; Green_fl_protein.
 DR Pfam; PF01353; GFP. 1.
 DR ProDom; PD013756; Green_fl_protein; 1.
 DR NON_TER 225
 FT SEQUENCE 225 AA; 25775 MW; 52DE2F716D083524 CRC64;

Query Match 20.1%; Score 255.5; DB 5; Length 225;
 Best Local Similarity 25.8%; Pred. No. 3.2e-13;
 Matches 65; Conservative 49; Mismatches 69; Indels 69; Gaps 9;

QY 10 AVVPIVLELDGVDYNGHKSFSVSGEGSDVYTGKTLKLF-ICTYGLPYWPPTLVTFESYGV 68
 Db 6 SVAKIKRLMDGIYNGHKSFSVSGEGSDVYTGKTLKLF-ICTYGLPYWPPTLVTFESYGV 65
 QY 69 QCFSRYPDHMKRHFESKAMPBGVYOQRTIFFKDDGNYKTRAEVKEGGDTLVNRIELKGI 128
 Db 66 RVETKYPKIDIP--DYFKQSPPEGSYSWERSMTFEDQGVCTVSDIKLEGDCEFFYEIRFGYV 123
 QY 129 DFEKEDGNILGHK-LEYNSHNYIMADKOKNGI-----KYNFKIRHNI 170
 Db 124 NFSSSGPVYMKTKLKEPSTENMY-----RDGVLLGDVSRITLLBEGDKHRCNFRSTYG 178
 QY 171 IEDGSVOLADYGOONPILDGVPVLPDNNHLSQSAISKDPNEKRDHMLLEFVTAAGIT 230
 Db 179 AKKG-----VVLPEYHFV-----DHRIRI-----LS 199
 QY 231 HGMD---ELTK 238
 Db 200 HDKDYNTVEYVE 211

RESULT 15

0963F5 PRELIMINARY; PRT; 225 AA.

AC 0963F5; PRELIMINARY; PRT; 225 AA.
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Green fluorescent protein.
 OS Montastraea cavernosa (great star coral).

OC Eukaryota; Metazoa; Chordata; Anthozoa; Zoantharia; Scleractinia;
 OC Faviina; Faviidae; Montastraea.
 OX NCBI_TaxID=63558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lesser M.P., Barry T.M., Mazel C., Matz M.V., Lukyanov S.A.,
 RA Falkowski P., Gorbunov M., Kolber Z.;
 RT "Green fluorescent proteins in Caribbean Scleractinian corals."
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF384683; AAK62982.2;
 DR InterPro; IPR000786; Green_fl_protein.
 DR Pfam; PF01353; GFP. 1.
 DR ProDom; PD013756; Green_fl_protein; 1.
 DR SEQUENCE 225 AA; 25847 MW; 77DE7D7C616929AF CRC64;

Query Match 19.3%; Score 245; DB 5; Length 225;
 Best Local Similarity 28.2%; Pred. No. 2.3e-12;
 Matches 58; Conservative 47; Mismatches 89; Indels 12; Gaps 6;

QY 11 VVPIVLELDGVDYNGHKSFSVSGEGSDVYTGKTLKLFICTTG-KLPVWPPTLVTFESYGV 69
 Db 7 IMEIKLMGVVNGHKSFSVSGEGSDVYTGKTLKLFICTTG-KLPVWPPTLVTFESYGV 66
 QY 70 CFSRYPDHMKRHFESKAMPBGVYOQRTIFFKDDGNYKTRAEVKEGGDTLVNRIELKGI 129
 Db 67 VETKYPKIDIP--DYFKQSPPEGSYSWERSMTFEDQGVCTVSDIKLEGDCEFFYEIRFGYV 124
 QY 130 FKEDGNILGHK-LEYNSHNYIMADKOKNGI-KYNFKIRHNIEDGSVOLADYGOONP 187
 Db 125 FPNPGPVYMKTKLKEPSTENMY-----RDGVLLGDVSRITLLBEGDKHRCNFRSTYG 177
 QY 188 ILDGVPVLPDNNHLSQSAISKDPNE 213
 Db 178 KAKKRVQLPDHYHVDHRIELSDND 203

Search completed: July 11, 2003, 17:58:45
 Job time : 83 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 17:51:18 ; Search time 23 Seconds

(without alignments)
429.190 Million cell updates/sec

Title: US-10-071-976-2

Perfect score: 1270
Sequence: 1 MSKGELEFTRVAVPILVELDG.....VLEFVTAAGITGMDELYK 238

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1242	97.8	238	1	GFP_AEQVI
2	100.5	7.9	879	1	STL_XYLEA
3	99	7.8	336	1	YD48_MENJA
4	90	7.1	795	1	D152_HAEIN
5	90	7.1	797	1	D151_HAEIN
6	89.5	7.0	793	1	MT04_HELPE
7	89	7.0	793	1	D153_HAEIN
8	88.5	7.0	312	1	TRXB_CHLMU
9	87.5	6.9	861	1	SYL_HAEIN
10	87	6.9	2334	1	WAPA_BACSU
11	86	6.8	357	1	TRMA_CAMJE
12	85.5	6.7	752	1	NEC1_RAT
13	85.5	6.7	886	1	TRH3_MESAU
14	84.5	6.7	312	1	TRXB_CHLMU
15	84.5	6.7	538	1	GRBE_RAT
16	82.5	6.5	613	1	PEPF_MYCPU
17	82.5	6.5	1164	1	BAG_STRAG
18	82	6.5	1433	1	DPO3_BACHD
19	81	6.4	836	1	DPO1_HPBUD
20	81	6.4	953	1	YMY7_YEAST
21	80.5	6.3	355	1	PLK_CHICK
22	80.5	6.3	504	1	MPPA_SOLTU
23	80.5	6.3	2222	1	DPOE_YEAST
24	80.5	6.3	6359	1	BACC_BACLI
25	80	6.3	774	1	AMT2_SCHPO
26	79.5	6.3	533	1	NIFD_CLOPA
27	79.5	6.3	887	1	TRH3_RAT
28	79.5	6.3	1224	1	COPA_HUMAN
29	79	6.2	339	1	PMAR_CANGA
30	79	6.2	354	1	PLK_HORSE
31	79	6.2	3135	1	S230_PLAFO
32	78	6.1	217	1	EXPI_ERWCA
33	78	6.1	520	1	CC3_YEAST

34	78	6.1	664	1	SYM_BACSU	P27465 bacillus su
35	77.5	6.1	269	1	PENK_MOUSE	P22005 mus musculu
36	77.5	6.1	461	1	PSBC_CVAPA	P48104 cyanophora
37	77.5	6.1	540	1	NSA_MCGPN	P75591 mycoplasma
38	77.5	6.1	682	1	PRC_ECOLI	P23865 escherichia
39	77.5	6.1	693	1	TGM3_HUMAN	Q08188 homo sapien
40	77.5	6.1	860	1	SYL_ECOLI	P07813 escherichia
41	77.5	6.1	1228	1	NARG_BACSU	P42175 bacillus su
42	77.5	6.1	1666	1	CO3_CAVPO	P12387 cavia porce
43	77	6.1	438	1	SHPI_STAPH	Q08002 staphylococ
44	77	6.1	753	1	NEC1_MOUSE	P21662 mus musculu
45	76.5	6.0	220	1	TTM2_CAVPO	Q9WUC6 cavia porce

ALIGNMENTS

RESULT 1
ID GFP_AEQVI STANDARD; PRT; 238 AA.
AC P42212; 017104;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxId=6100;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92175527; PubMed=1347277;
RA Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,
RA Cormier M.J.;
RT "Primary structure of the Aequorea victoria green-fluorescent
RT protein.";
RL Gene 111:229-233(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94185810; PubMed=8137953;
RA Inouye S., Tsuji F.I.;
RT "Aequorea green fluorescent protein. Expression of the gene and
RT fluorescence characteristics of the recombinant protein.";
RL FEBS Lett. 341:277-280(1994).
RN [3]
RP CHROMOPHORE.
RX MEDLINE=93192221; PubMed=8448132;
RA Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;
RT "Chemical structure of the hexapeptide chromophore of the Aequorea
RT green-fluorescent protein.";
RL Biochemistry 32:1212-1218(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=96355665; PubMed=8703075;
RA Ormeo M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,
RA Remington S.J.;
RT "Crystal structure of the Aequorea victoria green fluorescent
RT protein.";
RL Science 273:1392-1395(1996).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=98294543; PubMed=9631087;
RA Yang F., Moss L.G., Phillips G.N. Jr.;
RT "The molecular structure of green fluorescent protein.";
RL Nat. Biotechnol. 14:1246-1251(1996).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.
RX MEDLINE=98455509; PubMed=9782051;
RA Wachtler R.M., Elsliger M.A., Kallio K., Hanson G.T., Remington S.J.;
RT "Structural basis of spectral shifts in the yellow-emission variants
RT of green fluorescent protein.";
RL structure 6:1267-1277(1998).

RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=99238303; Pubmed=10220315;
 RA Eisliger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;
 RT "Structural and spectral response of green fluorescent protein
 RT variants to changes in pH."
 RL Biochemistry 38:5296-5301(1999).
 CC -1- FUNCTION: ENERGY-TRANSFER ACCEPTOR. ITS ROLE IS TO TRANSDUCE THE
 CC BLUE CHEMILUMINESCENCE OF THE PROTEIN AECORIN INTO GREEN
 CC FLUORESCENT LIGHT BY ENERGY TRANSFER. FLUORESCES IN VIVO UPON
 CC RECEIVING ENERGY FROM THE CA(2+)-ACTIVATED PHOTOPROTEIN AECORIN.
 CC ABSORBS LIGHT MAXIMALLY AT 395 NM AND EXHIBITS A SMALLER
 CC ABSORPTION PEAK AT 470 NM. THE FLUORESCENCE EMISSION SPECTRUM
 CC PEAKS AT 509 NM WITH A SHOULDER AT 540 NM.
 CC -1- SUBUNIT: MONOMER.
 CC -1- TISSUE SPECIFICITY: PHOTOCYTES.
 CC -1- PTM: CONTAINS A COVALENTLY ATTACHED CHROMOPHORE, WHICH IS COMPOSED
 CC OF MODIFIED AMINO ACID RESIDUES. THE CHROMOPHORE IS FORMED UPON
 CC CYCLIZATION OF THE RESIDUES SER-DEHYDROTYR-GLY.
 CC -1- BIOTECNOLOGY: Has become a useful and ubiquitous tool for making
 CC chimeric proteins of GFP linked to other proteins where it
 CC functions as a fluorescent protein tag. GFP tolerates N- and C-
 CC terminal fusion to a broad variety of proteins. It has been
 CC expressed in bacteria, yeast, slime mold, plants, drosophila,
 CC zebrafish, and in mammalian cells. As a noninvasive fluorescent
 CC marker in living cells, it allows for a wide range of applications
 CC where it may function as a cell lineage tracer, reporter of gene
 CC expression, or as a measure of protein-protein interactions.
 CC -1- DATABASE: NAME=protein Spotlight;
 CC NOTE=Issue 11 of June 2001;
 CC WWW=ftp://www.expasy.org/spotlight/articles/spt1011.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M62654; AAA27722.1; -;
 DR EMBL: M62653; AAA27721.1; -;
 DR EMBL: L29345; AAA58246.1; -;
 DR PDB: 1GFL; 11-JAN-97.
 DR PDB: 1EMA; 08-NOV-96.
 DR PDB: 1EMA; 16-JUN-97.
 DR PDB: 1EMC; 20-AUG-97.
 DR PDB: 2EMD; 20-AUG-97.
 DR PDB: 1EMF; 20-AUG-97.
 DR PDB: 1EMF; 20-AUG-97.
 DR PDB: 1EMG; 12-MAY-99.
 DR PDB: 1EMK; 20-AUG-97.
 DR PDB: 1EML; 20-AUG-97.
 DR PDB: 1EMM; 20-AUG-97.
 DR PDB: 2EMN; 20-AUG-97.
 DR PDB: 2EMO; 20-AUG-97.
 DR PDB: 1BFP; 07-JUL-97.
 DR PDB: 1YFP; 28-OCT-98.
 DR InterPro: IPR000786; Green_fl_protein.
 DR Pfam: PF01353; GFP; 1.
 DR PRINTS: PR01229; GFP; 1.
 DR PRODOM: PD013756; Green_fl_protein; 1.
 KW Luminescence; 3D-structure.
 FT SITE 65 67 MODIFIED TO FORM THE CHROMOPHORE.
 FT VARIANT 100 100 F -> Y.
 FT VARIANT 108 108 T -> S.
 FT VARIANT 141 141 L -> M.
 FT VARIANT 219 219 V -> I.
 FT VARIANT 25 25 H -> Q (IN REF. 2).
 FT VARIANT 157 157 Q -> P (IN REF. 2).
 FT VARIANT 172 172 E -> K (IN REF. 2).
 SO SEQUENCE 238 AA; 26886 MW; EA5A6F21FBFB6E05 CRC64;

Query Match 97.8%; Score 1242; DB 1; Length 238;
 Best Local Similarity 97.5%; Pred. No. 4.4e-98;
 Matches 233; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

 QY 1 MSKGEELFVAVPLVLDGVDGNGHKSVSGBGBDYTKLTKFKCTTGKLPVPTL 60
 DB 1 MSKGEELFVAVPLVLDGVDGNGHKSVSGBGBDYTKLTKFKCTTGKLPVPTL 60
 QY 61 VTFFSVGVQCFSRPDHMKQHDFFKSAMPGYVOERTIFFKDDGNYTRAEVFEQTLV 120
 DB 61 VTFFSVGVQCFSRPDHMKQHDFFKSAMPGYVOERTIFFKDDGNYTRAEVFEQTLV 120
 QY 121 NRTELKIDKREDGNILGHLREYNVNSHNYIMADKQKNGIKVFKIRNIEDGSVOLAD 180
 DB 121 NRTELKIDKREDGNILGHLREYNVNSHNYIMADKQKNGIKVFKIRNIEDGSVOLAD 180
 QY 181 YVOQNPFLGPPVLPNPNHLSQSALSQPNKRDHMLLEVTAAGTGHGDELYK 238
 DB 181 YVOQNPFLGPPVLPNPNHLSQSALSQPNKRDHMLLEVTAAGTGHGDELYK 238
 DB 181 HYQNPFGIDGPPVLPNPNHLSQSALSQPNKRDHMLLEVTAAGTGHGDELYK 238

 RESULT 2
 SYL_XYLFA STANDARD; PRT; 879 AA.
 AC 09PB68;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
 GN LEUS OR XF2176.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 CC Xylella.
 CC NCBI_TaxID=2371.
 OX
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9a5c;
 RX MEDLINE=203657717; Pubmed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.R.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marinho C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsushima A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.F.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Pelxoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pessegueiro J.B.,
 RA Quaggio R.B., Roberto P.A.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Savasati H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Tsunako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zaitz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa."
 RL Nature 406:151-159(2000).
 CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
 CC diphosphate + L-leucyl-tRNA(Leu).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL--TRNA SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

[illegible]

```

RT Rd.":
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: U32773; AAC22575.1; -.
DR TIGR: H10917; -.
DR InterPro: IPR00184; Bac.surfAg_D15.
DR Pfam: PF01103; Bac.surfAg_1.
KW Antigen; Outer membrane; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 795 POTENTIAL.
SQ SEQUENCE 795 AA; 87478 MW; B85691FC22E6ED4 CRC64;

Query Match 7.1%; Score 90; DB 1; Length 795;
Best Local Similarity 20.2%; Pred. No. 4.8;
Matches 52; Conservative 36; Mismatches 96; Indels 74; Gaps 12;

QY 22 VNGHK-----FSVSGEGEDVYTGKLTLEKICTGKLPVWPPLVTTFSYGVOCFSRYP 75
DB 402 INGSNDEVDVYKVKERNTGSING-----IGYRSGISYQ 438
QY 76 DHKRNDF-----FKSAMPGEYVOQ-----RTIEFKDGNYYTRA 110
DB 439 ASVKQDNFLGTGAASVAGTKNDYGVSNLGYTEPYTKDGVSLGWNVFENYDNSKSDT 498
QY 111 EVKEEGDTLVNRIELKIDKEKDGNI--LGH-----KLEYNNSHWYIMADQK- 158
DB 499 SSNKKRTTYSNVTL-GFPYNNNSYVVGIGHYNNKISNPALEN--RLUYIOSMKFKG 554
QY 159 NGIKVN-FKIRHNIEDSGVOLADYQOQNPFI-LDGPVLLP--DNHYLSTOSALSKDENEK 214
DB 555 NGIKTNDPFSFGMNNYSILNRGYEPTKGVKASLGGRVTIPGSDNKKYKLSADVOGFYPLD 614
QY 215 RDHNVLEFVTAAGITHG 232
DB 615 RDHLWVVSARASAGYANG 632

RESULT 5
D151_HAEIN STANDARD; PRT; 797 AA.
ID D151_HAEIN
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-A9) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC Haemophilus.
OC NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype B;
RA MEDLINE=95255676; PubMed=7737523;
RA Flack F.S., Loomore S., Chong P., Thomas W.R.;
RT "The sequencing of the 80-kDa D15 protective surface antigen of
RT Haemophilus influenzae.";
RL Gene 136:97-99(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Mina, and Eagan / Serotype B;
RA MEDLINE=97427952; PubMed=9286140;
RA Loomore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,

```

```

RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease.";
RL Infect. Immun. 65:3701-3707(1997).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: U13961; AAA85645.1; -.
DR EMBL: U60832; AAB61974.1; -.
DR EMBL: U60833; AAB61976.1; -.
DR InterPro: IPR00184; Bac.surfAg_D15.
DR Pfam: PF01103; Bac.surfAg_1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 19
FT CHAIN 20 797 POTENTIAL.
SQ SEQUENCE 797 AA; 87675 MW; 2F93DE538696AFLB CRC64;

Query Match 7.1%; Score 90; DB 1; Length 797;
Best Local Similarity 20.2%; Pred. No. 4.8;
Matches 52; Conservative 36; Mismatches 96; Indels 74; Gaps 12;

QY 22 VNGHK-----FSVSGEGEDVYTGKLTLEKICTGKLPVWPPLVTTFSYGVOCFSRYP 75
DB 402 INGSNDEVDVYKVKERNTGSING-----IGYRSGISYQ 438
QY 76 DHKRNDF-----FKSAMPGEYVOQ-----RTIEFKDGNYYTRA 110
DB 439 ASVKQDNFLGTGAASVAGTKNDYGVSNLGYTEPYTKDGVSLGWNVFENYDNSKSDT 498
QY 111 EVKEEGDTLVNRIELKIDKEKDGNI--LGH-----KLEYNNSHWYIMADQK- 158
DB 499 SSNKKRTTYSNVTL-GFPYNNNSYVVGIGHYNNKISNPALEN--RLUYIOSMKFKG 554
QY 159 NGIKVN-FKIRHNIEDSGVOLADYQOQNPFI-LDGPVLLP--DNHYLSTOSALSKDENEK 214
DB 555 NGIKTNDPFSFGMNNYSILNRGYEPTKGVKASLGGRVTIPGSDNKKYKLSADVOGFYPLD 614
QY 215 RDHNVLEFVTAAGITHG 232
DB 615 RDHLWVVSARASAGYANG 632

RESULT 6
MT04_HELPY STANDARD; PRT; 393 AA.
ID MT04_HELPY
AC Q25443;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical methyltransferase HP0747 (Ec 2.1.1.-).
GN HP0747.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OC NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RA MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khairi H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,

```

RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Meldman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.:
 RT "The complete genome sequence of the gastric pathogen *Helicobacter*
 RT *pylori*.";
 RL Nature 388:539-547(1997).
 CC -1- FUNCTION: PROBABLE METHYLTRANSFERASE.
 CC -1- SIMILARITY: BELONGS TO THE UPF0155 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AE000587; AAD07796.1; -
 DR TIGR; HP0747; -
 DR InterPro: IPR004395; Cons_hypoth91.
 DR InterPro: IPR003358; Methyltransf_4.
 DR InterPro: IPR000051; SAM_bind.
 DR Pfam; PF02390; Methyltransf_4; 1.
 DR TIGRFAMs; TIGR00091; Cons_hypoth91; 1.
 KW Hypothetical protein; Transferase; Methyltransferase;
 KW Complete proteome.
 SQ SEQUENCE 393 AA; 45774 MW; C376F02D33F71D1 CAC64;

Query Match 7.0%; Score 89.5; DB 1; Length 393;
 Best Local Similarity 20.8%; Pred. No. 2.3;
 Matches 51; Conservative 42; Mismatches 77; Indels 75; Gaps 15;

QY 34 EGDYTGKLTGK-----FICTGKLPVMPPLVLTSTYGVQCSRPDMKRR-- 81
 DB 174 QGD---GRVLVESPNERCEKIFV---HFPVPW-----MEKKRR 207
 QY 82 ----DFKFSAM-----PEGYVOORT---IFFKDD-----GNYKTRAEVFEED-TLVNRIE 124
 DB 208 VLSKFLNALRVKLRGFLRLRTDLSLFEEDSLKLNKQCELEIKNAQIPVSKYE 267
 QY 125 LKGIQDEKD-GNIGKHLKLVN---YNSHNYIMADK---QRNGIKVNEKIRHNIEDG-SV 176
 DB 268 ARWKKLKKDIYDLRIYSLEWNETPEFNHAFSEFETIRISKSVGTILTKKIIQEGYFV 327
 QY 177 QLADYVOQNTPII-----DGR-----VLLPDNHYLSTQSALSDPBEKDDHWLLEFV 224
 DB 328 HVCNHYENKGFVLVLSKGFDPVPLRFLVLTENOIFYLNKSPKTLNNHKAHLLONIL 387

QY 225 TAAGI 229
 DB 388 SQKGI 392

RESULT 7
 ID D153_HAEIN STANDARD; PRT; 793 AA.
 AC 032629;

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
 DE (D-15-Ag) (Outer membrane protein D15).
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PAK 12085;
 RX MEDLINE=97427952; PubMed=9284140;
 RA Loosmore S.M., Tang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
 RA Klein M.H.;

RT "Outer membrane protein D15 is conserved among *Haemophilus influenzae*
 RT species and may represent a universal protective antigen against
 RT invasive disease."
 RL Infect. Immun. 65:3701-3707(1997).
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; U60834; AAB61977.1; -
 DR InterPro: IPR000184; Bac_surfAg_D15.
 DR Pfam; PF01103; Bac_surface_Ag; 1.
 KW Antigen; Outer membrane; Signal.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
 SQ SEQUENCE 793 AA; 87511 MW; 51BFB203680A1A4 CAC64;

Query Match 7.0%; Score 89; DB 1; Length 793;
 Best Local Similarity 20.5%; Pred. No. 5.8;
 Matches 53; Conservative 34; Mismatches 97; Indels 74; Gaps 12;

QY 22 VNGHK-----FVSGBEGDYTGKLTGKLTCTTGKLPVMPPLVLTSTYGVQCSKYP 75
 DB 402 INGSNDEVYVYKVERNTGSIING-----RTIFFKDDGKYKRA 110
 QY 76 DHMKRHF-----FKSAMPEGYVOO-----IGGTSGISYQ 438
 DB 439 TSIKODNLTGCAAVSINAGTNDYGTSVNLGTEPFTFKDQVSLAGNIFENYDNSKDT 498
 QY 111 EVKEGDTLVRIELKGIQDEKDNIT---LGH-----KLEYNNSHNYIMADKOK 158
 DB 499 SSNKRRTYGSNVTL-GFPVNENSYVGLGHTYKIKSNFALRYN---RNLYQSMKRFK 554
 QY 159 NGIKVY-FKIRHNIEDGSVQADLYQOQNTPI-LDGVPVLLP--DNHYLSTQSALSDPBEK 214
 DB 555 NGIKTNDDFSGFMWYNSLNGRYPTPKGVKASLGGRVYIIPSDMKYKYLSDVQGYPLD 614
 QY 215 RDHWLLEFYTAAGITHG 232
 DB 615 RDHRWVYSAKASAGYANG 632

RESULT 8

TRXB_CHLMU STANDARD; PRT; 312 AA.
 ID TRXB_CHLMU
 AC 09PKT7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thioresdoxin reductase (EC 1.6.4.5) (TRXR).
 GN TRXB OR TC0375.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=MOPn / N19g;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.R., Peterson J., Utterback T., Berry K., Baas S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of *Chlamydia trachomatis* MOPn and *Chlamydia*
 RT *pneumoniae* AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -1- CATALYTIC ACTIVITY: NADPH + oxidized thioresdoxin = NADP(+) +


```

CC reduced thioredoxin.
CC -1- COFACTOR: FAD.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC OXIDOREDUCTASES CLASS-II.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE002304; AAF39233.1; -.
DR HSSP: Q39243; IYDC.
DR TIGR: TC0375; -.
DR InterPro: IPR000759; Adnrdx_reductase.
DR InterPro: IPR001327; FAD_pyr_redox.
DR InterPro: IPR001100; Pyr_redox.
DR Pfam: PF00070; Pyr_redox_1.
DR PRINTS: PR00419; ADXRDASE.
DR PRINTS: PR00368; FADPRR.
DR PRINTS: PR00411; PNDRDASE1.
DR PRINTS: PR00469; PNDRDASE1.
DR ProDom: PD000139; FAD_pyr_redox; 1.
DR TIGRFAMs: TIGR01292; TRX_reduct; 1.
DR PROSITE: PS00573; PYRIDINE_REDOX_2; 1.
DR Redox-active center: Oxidoreductase; NADP; Flavoprotein; FAD;
KW Complete proteome.
FT NP_BIND 4 19 FAD (ADP PART) (PROBABLE).
FT DISULFID 138 141 REDOX-ACTIVE (BY SIMILARITY).
FT NP_BIND 152 166 NAD(P) (BY SIMILARITY).
FT NP_BIND 273 283 FAD (FLAVIN PART) (BY SIMILARITY).
FT SEQUENCE 312 AA; 33556 MW; 53824B207C2155D CRC64;
SQ
Query Match 7.0%; Score 88.5; DB 1; Length 312;
Best Local Similarity 21.1%; Pred. No. 2.1;
Matches 48; Conservative 29; Mismatches 81; Indels 69; Gaps 9;
QY 1 MSGGEELFVAVPIVLELDGVNGHKSVSSEGGSDVYTKLTKFCTTGKLPVPTL 60
DB 97 LKSGEDFTCDACIAT--GASAKRLSLPGAGDNEFMORGVACAVC--DGASPI----- 147
QY 61 VTFESYGVCFSRYPDMKRHDFF-----KSAMEGVVQORTIFEKDGNVKTAEYKFE 115
DB 148 -----FRDRDLFVIGGDSALEAMFLTR-----YGRKRVYVHR 181
QY 116 GDPL-----VNRLEKLGIDFKEDGNILGHKLEYNNSHNYTIMADKQNGIKVNFRIH 169
DB 182 RDLTKASKAMVNAQA-----ANEKIVFLMNSVYVKILGDSLVRSI-----DIFN 225
QY 170 NIDSGSYQL-----ADYQOQNTPLDGPVLLPDPNHYLSTQSLK 209
DB 226 NVEKTYTMAAGVFAIGHQPTAFLGQSLSDENGYITTEKSSR 272

```

RESULT 9
SYL_HAEIN
ID SYL_HAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUCS OR HI0921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxId=727;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KM20 / ATCC 51907;
RX MEDLINE-95350630; PubMed-7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kesteven A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInerney K., Sutton G., Fitznugh W., Fields C.A., Goodyne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U33774; AAC22581.1; -.
DR TIGR: HI0921; -.
DR InterPro: IPR002302; Leu-tRNA-synt.1a.
DR InterPro: IPR002300; tRNA-synt.1a.
DR InterPro: IPR001412; tRNA-synt.1.
DR Pfam: PF00133; tRNA-synt.1; 1.
DR PRINTS: PR00985; TRNASYNTLEU.
DR TIGRFAMs: TIGR00396; LeuS_bact; 1.
DR PROSITE: PS00178; AA_tRNA_LIGASE_1; 1.
DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
FT SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FB7 CRC64;
SQ
Query Match 6.9%; Score 87.5; DB 1; Length 861;
Best Local Similarity 24.3%; Pred. No. 8.6;
Matches 45; Conservative 25; Mismatches 68; Indels 47; Gaps 9;
QY 49 TTGKLPVWPPTLVTFESYGVCFSRYPDMKRHDFFKSAMEGVVQORTIFEKDGNVKT 108
DB 314 TGDKLPI-WVANFVLMHYGTGAWAAPAHQORD--FEFAQKYSTPIKQVIAPIADERIDL 370
QY 109 RAEVKPEGDLVNRLEKLGIDFKEDGNILGHKLEYNNSHNYTIMADK--QNGI---KVN 164
DB 371 TKQAFVEHGKLVNSDEFGNF--DGAFGN-----LADKLEKLVGKRQVN 414
QY 165 FKIRH-----NIDSGSYQLADYQOQNTPLDGPVLLPDPNHYLSTQSL 207
DB 415 YRLDQWVSQRQRYWGAPIPMLTLENGDVYRA-----PREDLPILPEVYWDGVKSPI 467

```

RESULT 10
WAPA_BACSU
ID WAPA_BACSU STANDARD; PRT; 2334 AA.
AC Q07833;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Wall-associated protein precursor.
 GN WAPA OR NI7G.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 NC NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=93302506; PubMed=8316082;
 RA Foster S.J.;
 RT "Molecular analysis of three major wall-associated proteins of
 Bacillus subtilis 168: evidence for processing of the product of a
 gene encoding a 258 kDa precursor two-domain ligand-binding
 protein."
 RL Mol. Microbiol. 8:299-310(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=95219088; PubMed=7704263;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
 genome containing the hut and wapa loci."
 RL Microbiology 141:337-343(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=97124196; PubMed=8969509;
 RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
 Mitsu Y., Fujita Y.;
 RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
 containing the lic and cel loci, and creation of a 177 kb contig
 covering the gut-saxx region."
 RL Microbiology 142:3113-3123(1996).
 CC -1- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
 MOTILITY, SECRETION OR DIFFERENTIATION.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
 INTO THE MEDIUM.
 CC -1- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS: THE N-TERMINUS, HAS THREE
 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
 C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
 MOTIF REPEATED 31 TIMES.
 CC -1- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
 SIMILARITY TO THE REPEAT IN E. COLI RBS GROUP OF PROTEINS (RBSA-D).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L05634; AAA22883.1; -;
 DR EMBL; D31856; BAA0656.1; -;
 DR EMBL; D29985; BAA06260.1; -;
 DR EMBL; D83026; BAA11683.1; -;
 DR EMBL; Z99124; CAB15959.1; -;
 DR PIR; S32920; S32920.
 DR Subtilist; BG10797; WAPA.
 DR InterPro; IPR003305; CBM_Cenc.
 KW Pfam; PF02018; CBM_4_9; 1.
 KW Cell wall; Repeat; Complete proteome.
 FT SIGNAL 1 28 OR 32 (POTENTIAL).
 FT CHAIN 29 2334 WALL-ASSOCIATED PROTEIN.
 FT DOMAIN 504 869 3 X 101 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 504 605 1-1.
 FT REPEAT 636 736 1-2.
 FT REPEAT 769 869 1-3.
 FT DOMAIN 1021 2139 31 X 21 AA APPROXIMATE TANDEM REPEATS OF
 FT REPEAT 1021 1040 X(4)-G-X(4)-(YF)-X-D-X(2)-G-X(4).
 FT REPEAT 1042 1061 2-1.
 FT REPEAT 2-2.

FT REPEAT 1063 1082 2-3.
 FT REPEAT 1083 1102 2-4.
 FT REPEAT 1109 1128 2-5.
 FT REPEAT 1129 1148 2-6.
 FT REPEAT 1150 1169 2-7.
 FT REPEAT 1174 1193 2-8.
 FT REPEAT 1199 1218 2-9.
 FT REPEAT 1219 1238 2-10.
 FT REPEAT 1249 1268 2-11.
 FT REPEAT 1267 1286 2-12.
 FT REPEAT 1290 1309 2-13.
 FT REPEAT 1311 1330 2-14.
 FT REPEAT 1332 1351 2-15.
 FT REPEAT 1373 1392 2-16.
 FT REPEAT 1414 1433 2-17.
 FT REPEAT 1454 1473 2-18.
 FT REPEAT 1494 1513 2-19.
 FT REPEAT 1534 1553 2-20.
 FT REPEAT 1574 1593 2-21.
 FT REPEAT 1614 1633 2-22.
 FT REPEAT 1654 1673 2-23.
 FT REPEAT 1694 1713 2-24 (APPROXIMATE).
 FT REPEAT 1734 1753 2-25.
 FT REPEAT 1774 1793 2-26.
 FT REPEAT 1814 1833 2-27.
 FT REPEAT 1854 1873 2-28.
 FT REPEAT 1894 1913 2-29.
 FT REPEAT 1934 1953 2-30.
 FT REPEAT 1974 1993 2-31.
 FT REPEAT 2014 2033 2-32.
 FT REPEAT 2054 2073 2-33.
 FT REPEAT 2094 2113 2-34.
 FT REPEAT 2134 2153 2-35.
 FT REPEAT 2174 2193 2-36.
 FT REPEAT 2214 2233 2-37.
 FT REPEAT 2254 2273 2-38.
 FT REPEAT 2294 2313 2-39.
 FT REPEAT 2334 2353 2-40.
 FT REPEAT 2374 2393 2-41.
 FT REPEAT 2414 2433 2-42.
 FT REPEAT 2454 2473 2-43.
 FT REPEAT 2494 2513 2-44.
 FT REPEAT 2534 2553 2-45.
 FT REPEAT 2574 2593 2-46.
 FT REPEAT 2614 2633 2-47.
 FT REPEAT 2654 2673 2-48.
 FT REPEAT 2694 2713 2-49.
 FT REPEAT 2734 2753 2-50.
 FT REPEAT 2774 2793 2-51.
 FT REPEAT 2814 2833 2-52.
 FT REPEAT 2854 2873 2-53.
 FT REPEAT 2894 2913 2-54.
 FT REPEAT 2934 2953 2-55.
 FT REPEAT 2974 2993 2-56.
 FT REPEAT 3014 3033 2-57.
 FT REPEAT 3054 3073 2-58.
 FT REPEAT 3094 3113 2-59.
 FT REPEAT 3134 3153 2-60.
 FT REPEAT 3174 3193 2-61.
 FT REPEAT 3214 3233 2-62.
 FT REPEAT 3254 3273 2-63.
 FT REPEAT 3294 3313 2-64.
 FT REPEAT 3334 3353 2-65.
 FT REPEAT 3374 3393 2-66.
 FT REPEAT 3414 3433 2-67.
 FT REPEAT 3454 3473 2-68.
 FT REPEAT 3494 3513 2-69.
 FT REPEAT 3534 3553 2-70.
 FT REPEAT 3574 3593 2-71.
 FT REPEAT 3614 3633 2-72.
 FT REPEAT 3654 3673 2-73.
 FT REPEAT 3694 3713 2-74.
 FT REPEAT 3734 3753 2-75.
 FT REPEAT 3774 3793 2-76.
 FT REPEAT 3814 3833 2-77.
 FT REPEAT 3854 3873 2-78.
 FT REPEAT 3894 3913 2-79.
 FT REPEAT 3934 3953 2-80.
 FT REPEAT 3974 3993 2-81.
 FT REPEAT 4014 4033 2-82.
 FT REPEAT 4054 4073 2-83.
 FT REPEAT 4094 4113 2-84.
 FT REPEAT 4134 4153 2-85.
 FT REPEAT 4174 4193 2-86.
 FT REPEAT 4214 4233 2-87.
 FT REPEAT 4254 4273 2-88.
 FT REPEAT 4294 4313 2-89.
 FT REPEAT 4334 4353 2-90.
 FT REPEAT 4374 4393 2-91.
 FT REPEAT 4414 4433 2-92.
 FT REPEAT 4454 4473 2-93.
 FT REPEAT 4494 4513 2-94.
 FT REPEAT 4534 4553 2-95.
 FT REPEAT 4574 4593 2-96.
 FT REPEAT 4614 4633 2-97.
 FT REPEAT 4654 4673 2-98.
 FT REPEAT 4694 4713 2-99.
 FT REPEAT 4734 4753 2-100.
 FT REPEAT 4774 4793 2-101.
 FT REPEAT 4814 4833 2-102.
 FT REPEAT 4854 4873 2-103.
 FT REPEAT 4894 4913 2-104.
 FT REPEAT 4934 4953 2-105.
 FT REPEAT 4974 4993 2-106.
 FT REPEAT 5014 5033 2-107.
 FT REPEAT 5054 5073 2-108.
 FT REPEAT 5094 5113 2-109.
 FT REPEAT 5134 5153 2-110.
 FT REPEAT 5174 5193 2-111.
 FT REPEAT 5214 5233 2-112.
 FT REPEAT 5254 5273 2-113.
 FT REPEAT 5294 5313 2-114.
 FT REPEAT 5334 5353 2-115.
 FT REPEAT 5374 5393 2-116.
 FT REPEAT 5414 5433 2-117.
 FT REPEAT 5454 5473 2-118.
 FT REPEAT 5494 5513 2-119.
 FT REPEAT 5534 5553 2-120.
 FT REPEAT 5574 5593 2-121.
 FT REPEAT 5614 5633 2-122.
 FT REPEAT 5654 5673 2-123.
 FT REPEAT 5694 5713 2-124.
 FT REPEAT 5734 5753 2-125.
 FT REPEAT 5774 5793 2-126.
 FT REPEAT 5814 5833 2-127.
 FT REPEAT 5854 5873 2-128.
 FT REPEAT 5894 5913 2-129.
 FT REPEAT 5934 5953 2-130.
 FT REPEAT 5974 5993 2-131.
 FT REPEAT 6014 6033 2-132.
 FT REPEAT 6054 6073 2-133.
 FT REPEAT 6094 6113 2-134.
 FT REPEAT 6134 6153 2-135.
 FT REPEAT 6174 6193 2-136.
 FT REPEAT 6214 6233 2-137.
 FT REPEAT 6254 6273 2-138.
 FT REPEAT 6294 6313 2-139.
 FT REPEAT 6334 6353 2-140.
 FT REPEAT 6374 6393 2-141.
 FT REPEAT 6414 6433 2-142.
 FT REPEAT 6454 6473 2-143.
 FT REPEAT 6494 6513 2-144.
 FT REPEAT 6534 6553 2-145.
 FT REPEAT 6574 6593 2-146.
 FT REPEAT 6614 6633 2-147.
 FT REPEAT 6654 6673 2-148.
 FT REPEAT 6694 6713 2-149.
 FT REPEAT 6734 6753 2-150.
 FT REPEAT 6774 6793 2-151.
 FT REPEAT 6814 6833 2-152.
 FT REPEAT 6854 6873 2-153.
 FT REPEAT 6894 6913 2-154.
 FT REPEAT 6934 6953 2-155.
 FT REPEAT 6974 6993 2-156.
 FT REPEAT 7014 7033 2-157.
 FT REPEAT 7054 7073 2-158.
 FT REPEAT 7094 7113 2-159.
 FT REPEAT 7134 7153 2-160.
 FT REPEAT 7174 7193 2-161.
 FT REPEAT 7214 7233 2-162.
 FT REPEAT 7254 7273 2-163.
 FT REPEAT 7294 7313 2-164.
 FT REPEAT 7334 7353 2-165.
 FT REPEAT 7374 7393 2-166.
 FT REPEAT 7414 7433 2-167.
 FT REPEAT 7454 7473 2-168.
 FT REPEAT 7494 7513 2-169.
 FT REPEAT 7534 7553 2-170.
 FT REPEAT 7574 7593 2-171.
 FT REPEAT 7614 7633 2-172.
 FT REPEAT 7654 7673 2-173.
 FT REPEAT 7694 7713 2-174.
 FT REPEAT 7734 7753 2-175.
 FT REPEAT 7774 7793 2-176.
 FT REPEAT 7814 7833 2-177.
 FT REPEAT 7854 7873 2-178.
 FT REPEAT 7894 7913 2-179.
 FT REPEAT 7934 7953 2-180.
 FT REPEAT 7974 7993 2-181.
 FT REPEAT 8014 8033 2-182.
 FT REPEAT 8054 8073 2-183.
 FT REPEAT 8094 8113 2-184.
 FT REPEAT 8134 8153 2-185.
 FT REPEAT 8174 8193 2-186.
 FT REPEAT 8214 8233 2-187.
 FT REPEAT 8254 8273 2-188.
 FT REPEAT 8294 8313 2-189.
 FT REPEAT 8334 8353 2-190.
 FT REPEAT 8374 8393 2-191.
 FT REPEAT 8414 8433 2-192.
 FT REPEAT 8454 8473 2-193.
 FT REPEAT 8494 8513 2-194.
 FT REPEAT 8534 8553 2-195.
 FT REPEAT 8574 8593 2-196.
 FT REPEAT 8614 8633 2-197.
 FT REPEAT 8654 8673 2-198.
 FT REPEAT 8694 8713 2-199.
 FT REPEAT 8734 8753 2-200.
 FT REPEAT 8774 8793 2-201.
 FT REPEAT 8814 8833 2-202.
 FT REPEAT 8854 8873 2-203.
 FT REPEAT 8894 8913 2-204.
 FT REPEAT 8934 8953 2-205.
 FT REPEAT 8974 8993 2-206.
 FT REPEAT 9014 9033 2-207.
 FT REPEAT 9054 9073 2-208.
 FT REPEAT 9094 9113 2-209.
 FT REPEAT 9134 9153 2-210.
 FT REPEAT 9174 9193 2-211.
 FT REPEAT 9214 9233 2-212.
 FT REPEAT 9254 9273 2-213.
 FT REPEAT 9294 9313 2-214.
 FT REPEAT 9334 9353 2-215.
 FT REPEAT 9374 9393 2-216.
 FT REPEAT 9414 9433 2-217.
 FT REPEAT 9454 9473 2-218.
 FT REPEAT 9494 9513 2-219.
 FT REPEAT 9534 9553 2-220.
 FT REPEAT 9574 9593 2-221.
 FT REPEAT 9614 9633 2-222.
 FT REPEAT 9654 9673 2-223.
 FT REPEAT 9694 9713 2-224.
 FT REPEAT 9734 9753 2-225.
 FT REPEAT 9774 9793 2-226.
 FT REPEAT 9814 9833 2-227.
 FT REPEAT 9854 9873 2-228.
 FT REPEAT 9894 9913 2-229.
 FT REPEAT 9934 9953 2-230.
 FT REPEAT 9974 9993 2-231.
 FT REPEAT 10014 10033 2-232.
 FT REPEAT 10054 10073 2-233.
 FT REPEAT 10094 10113 2-234.
 FT REPEAT 10134 10153 2-235.
 FT REPEAT 10174 10193 2-236.
 FT REPEAT 10214 10233 2-237.
 FT REPEAT 10254 10273 2-238.
 FT REPEAT 10294 10313 2-239.
 FT REPEAT 10334 10353 2-240.
 FT REPEAT 10374 10393 2-241.
 FT REPEAT 10414 10433 2-242.
 FT REPEAT 10454 10473 2-243.
 FT REPEAT 10494 10513 2-244.
 FT REPEAT 10534 10553 2-245.
 FT REPEAT 10574 10593 2-246.
 FT REPEAT 10614 10633 2-247.
 FT REPEAT 10654 10673 2-248.
 FT REPEAT 10694 10713 2-249.
 FT REPEAT 10734 10753 2-250.
 FT REPEAT 10774 10793 2-251.
 FT REPEAT 10814 10833 2-252.
 FT REPEAT 10854 10873 2-253.
 FT REPEAT 10894 10913 2-254.
 FT REPEAT 10934 10953 2-255.
 FT REPEAT 10974 10993 2-256.
 FT REPEAT 11014 11033 2-257.
 FT REPEAT 11054 11073 2-258.
 FT REPEAT 11094 11113 2-259.
 FT REPEAT 11134 11153 2-260.
 FT REPEAT 11174 11193 2-261.
 FT REPEAT 11214 11233 2-262.
 FT REPEAT 11254 11273 2-263.
 FT REPEAT 11294 11313 2-264.
 FT REPEAT 11334 11353 2-265.
 FT REPEAT 11374 11393 2-266.
 FT REPEAT 11414 11433 2-267.
 FT REPEAT 11454 11473 2-268.
 FT REPEAT 11494 11513 2-269.
 FT REPEAT 11534 11553 2-270.
 FT REPEAT 11574 11593 2-271.
 FT REPEAT 11614 11633 2-272.
 FT REPEAT 11654 11673 2-273.
 FT REPEAT 11694 11713 2-274.
 FT REPEAT 11734 11753 2-275.
 FT REPEAT 11774 11793 2-276.
 FT REPEAT 11814 11833 2-277.
 FT REPEAT 11854 11873 2-278.
 FT REPEAT 11894 11913 2-279.
 FT REPEAT 11934 11953 2-280.
 FT REPEAT 11974 11993 2-281.
 FT REPEAT 12014 12033 2-282.
 FT REPEAT 12054 12073 2-283.
 FT REPEAT 12094 12113 2-284.
 FT REPEAT 12134 12153 2-285.
 FT REPEAT 12174 12193 2-286.
 FT REPEAT 12214 12233 2-287.
 FT REPEAT 12254 12273 2-288.
 FT REPEAT 12294 12313 2-289.
 FT REPEAT 12334 12353 2-290.
 FT REPEAT 12374 12393 2-291.
 FT REPEAT 12414 12433 2-292.
 FT REPEAT 12454 12473 2-293.
 FT REPEAT 12494 12513 2-294.
 FT REPEAT 12534 12553 2-295.
 FT REPEAT 12574 12593 2-296.
 FT REPEAT 12614 12633 2-297.
 FT REPEAT 12654 12673 2-298.
 FT REPEAT 12694 12713 2-299.
 FT REPEAT 12734 12753 2-300.
 FT REPEAT 12774 12793 2-301.
 FT REPEAT 12814 12833 2-302.
 FT REPEAT 12854 12873 2-303.
 FT REPEAT 12894 12913 2-304.
 FT REPEAT 12934 12953 2-305.
 FT REPEAT 12974 12993 2-306.
 FT REPEAT 13014 13033 2-307.
 FT REPEAT 13054 13073 2-308.
 FT REPEAT 13094 13113 2-309.
 FT REPEAT 13134 13153 2-310.
 FT REPEAT 13174 13193 2-311.
 FT REPEAT 13214 13233 2-312.
 FT REPEAT 13254 13273 2-313.
 FT REPEAT 13294 13313 2-314.
 FT REPEAT 13334 13353 2-315.
 FT REPEAT 13374 13393 2-316.
 FT REPEAT 13414 13433 2-317.
 FT REPEAT 13454 13473 2-318.
 FT REPEAT 13494 13513 2-319.
 FT REPEAT 13534 13553 2-320.
 FT REPEAT 13574 13593 2-321.
 FT REPEAT 13614 13633 2-322.
 FT REPEAT 13654 13673 2-323.
 FT REPEAT 13694 13713 2-324.
 FT REPEAT 13734 13753 2-325.
 FT REPEAT 13774 13793 2-326.
 FT REPEAT 13814 13833 2-327.
 FT REPEAT 13854 13873 2-328.
 FT REPEAT 13894 13913 2-329.
 FT REPEAT 13934 13953 2-330.
 FT REPEAT 13974 13993 2-331.
 FT REPEAT 14014 14033 2-332.
 FT REPEAT 14054 14073 2-333.
 FT REPEAT 14094 14113 2-334.
 FT REPEAT 14134 14153 2-335.
 FT REPEAT 14174 14193 2-336.
 FT REPEAT 14214 14233 2-337.
 FT REPEAT 14254 14273 2-338.
 FT REPEAT 14294 14313 2-339.
 FT REPEAT 14334 14353 2-340.
 FT REPEAT 14374 14393 2-341.
 FT REPEAT 14414 14433 2-342.
 FT REPEAT 14454 14473 2-343.
 FT REPEAT 14494 14513 2-344.
 FT REPEAT 14534 14553 2-345.
 FT REPEAT 14574 14593 2-346.
 FT REPEAT 14614 14633 2-347.
 FT REPEAT 14654 14673 2-348.
 FT REPEAT 14694 14713 2-349.
 FT REPEAT 14734 14753 2-350.
 FT REPEAT 14774 14793 2-351.
 FT REPEAT 14814 14833 2-352.
 FT REPEAT 14854 14873 2-353.
 FT REPEAT 14894 14913 2-354.
 FT REPEAT 14934 14953 2-355.
 FT REPEAT 14974 14993 2-356.
 FT REPEAT 15014 15033 2-357.
 FT REPEAT 15054 15073 2-358.
 FT REPEAT 15094 15113 2-359.
 FT REPEAT 15134 15153 2-360.
 FT REPEAT 15174 15193 2-361.
 FT REPEAT 15214 15233 2-362.
 FT REPEAT 15254 15273 2-363.
 FT REPEAT 15294 15313 2-364.
 FT REPEAT 15334 15353 2-365.
 FT REPEAT 15374 15393 2-366.
 FT REPEAT 15414 15433 2-367.
 FT REPEAT 15454 15473 2-368.
 FT REPEAT 15494 15513 2-369.
 FT REPEAT 15534 15553 2-370.
 FT REPEAT 15574 15593 2-371.
 FT REPEAT 15614 15633 2-372.
 FT REPEAT 15654 15673 2-373.
 FT REPEAT 15694 15713 2-374.
 FT REPEAT 15734 15753 2-375.
 FT REPEAT 15774 15793 2-376.
 FT REPEAT 15814 15833 2-377.
 FT REPEAT 15854 15873 2-378.
 FT REPEAT 15894 15913 2-379.
 FT REPEAT 15934 15953 2-380.
 FT REPEAT 15974 15993 2-381.
 FT REPEAT 16014 16033 2-382.
 FT REPEAT 16054 16073 2-383.
 FT REPEAT 16094 16113 2-384.
 FT REPEAT 16134 16153 2-385.
 FT REPEAT 16174 16193 2-386.
 FT REPEAT 16214 16233 2-387.
 FT REPEAT 16254 16273 2-388.
 FT REPEAT 16294 16313 2-389.
 FT REPEAT 16334 16353 2-390.
 FT REPEAT 16374 16393 2-391.
 FT REPEAT 16414 16433 2-392.
 FT REPEAT 16454 16473 2-393.
 FT REPEAT 16494 16513 2-394.
 FT REPEAT 16534 16553 2-395.
 FT REPEAT 16574 16593 2-396.
 FT REPEAT 16614 16633 2-397.
 FT REPEAT 16654 16673 2-398.
 FT REPEAT 16694 16713 2-399.
 FT REPEAT 16734 16753 2-400.
 FT REPEAT 16774 16793 2-401.
 FT REPEAT 16814 16833 2-402.
 FT REPEAT 16854 16873 2-403.
 FT REPEAT 16894 16913 2-404.
 FT REPEAT 16934 16953 2-405.
 FT REPEAT 16974 16993 2-406.
 FT REPEAT 17014 17033 2-407.
 FT REPEAT 17054 17073 2-408.
 FT REPEAT 17094 17113 2-409.
 FT REPEAT 17134 17153 2-410.
 FT REPEAT 17174 17193 2-411.
 FT REPEAT 17214 17233 2-412.
 FT REPEAT 17254 17273 2-413.
 FT REPEAT 17294 17313 2-414.
 FT REPEAT 17334 17353 2-415.
 FT REPEAT 17374 17393 2-416.
 FT REPEAT 17414 17433 2-417.
 FT REPEAT 17454 17473 2-418.
 FT REPEAT 17494 17513 2-419.
 FT REPEAT 17534 17553 2-420.
 FT REPEAT 17574 17593 2-421.
 FT REPEAT 17614 17633 2-422.
 FT REPEAT 17654 17673 2-423.
 FT REPEAT 17694 17713 2-424.
 FT REPEAT 17734 17753 2-425.
 FT REPEAT 17774 17793 2-426.
 FT REPEAT 17814 17833 2-427.
 FT REPEAT 17854 17873 2-428.
 FT REPEAT 17894 17913 2-429.
 FT REPEAT 17934 17953 2-430.
 FT REPEAT 17974 17993 2-431.
 FT REPEAT 18014 18033 2-432.
 FT REPEAT 18054 18073 2-433.
 FT REPEAT 18094 18113 2-434.
 FT REPEAT 18134 18153

RA MEDLINE-20150912; PubMed-10688204;
 RA Parkhill J., Wren B.W., Mungall K., Kelley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
 RA Jagsels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rasthream M.A., Rutherford K.M., van Vleet A.H.M.,
 RA Whitehead S., Barrall B.G.,
 RT "The genome sequence of the food-borne pathogen *Campylobacter jejuni*
 RT reveals hypervariable sequences.";
 RL Nature 403:665-668(2000).
 CC -1- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position
 CC 54 (M-5-054) in all tRNA (By similarity).
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
 CC homocysteine + tRNA containing thymine.
 CC -1- SIMILARITY: BELONGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. TRMA
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL39076; CAB73096.1; -
 DR InterPro: IPR000051; SAM_Dbind.
 DR InterPro: IPR001566; TRMA.
 DR PROSITE: PS01230; TRMA_1; 1.
 DR PROSITE: PS01231; TRMA_2; FALSE_NEG.
 KM Transferrase; Methyltransferase; tRNA processing; Complete proteome.
 FT DOMAIN 207 213 S-ADENOSYLMETHIONINE BINDING (BY
 FT SIMILARITY).
 FT ACT_SITE 315 315 BY SIMILARITY.
 SQ SEQUENCE 357 AA; 42276 MW; CEC5328347CEEA97 CRC64;
 Query Match 6.8%; Score 86; DB 1; Length 357;
 Best Local Similarity 21.8%; Pred. No. 4; Indels 62; Gaps 6;
 Matches 36; Conservative 23; Mismatches 44;
 QY 79 KRHDFKSAPEGVQORTIFPKDGNKTRAEVKE--EGDTLV----- 120
 Db 14 EKHSFKYKFEFTYKDFKFLASDKHRTAEISFYHENDTLFYAMFDPKSKKYLIEY 73
 QY 121 -----NRIELK--GIDPKEDGNILGKLEYNYN----- 146
 Db 74 LDFADEKICAFMPRLLEYLRQDNKLEKLEGVETLTKQELSTILLYLHKNIEDIKSNLEN 133
 QY 147 -SHNVYIMADKOKNGIKVNFKIRH-----NIEGSOVLADYOOON 185
 Db 134 LSNILHNLJARSKKKILFKTENTLROTLMNIODRKI---FYEFN 174
 RESULT 12
 NECL_RAT STANDARD; PRT; 752 AA.
 AC P28840;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-CCV-2001 (Rel. 40, Last annotation update)
 DE Neuroendocrine convertase 1 precursor (EC 3.4.21.93) (NRC 1) (PCL)
 DE (Prothormone convertase 1) (Prothormone convertase 1).
 GN PCSK1 OR NECL OR NEC-1 OR BDP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92168040; PubMed-1791845;
 RA Bloomquist B.T., Ripper B.A., Mains R.E.;
 RT "Prothormone-converting enzymes: regulation and evaluation of function
 RT using antisense RNA.";
 RL Mol. Endocrinol. 5:2014-2024(1991).

RA [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92063860; PubMed-1954888;
 RA Hakes D.J., Birch N.P., Mezey A., Dixon J.E.;
 RT "Isolation of two complementary deoxyribonucleic acid clones from a
 RT rat insulinoma cell line based on similarities to Kex2 and furin
 RT sequences and the specific localization of each transcript to
 RT endocrine and neuroendocrine tissues in rats.";
 RL Endocrinology 129:3053-3063(1991).
 CC -1- FUNCTION: INVOLVED IN THE PROCESSING OF HORMONE AND OTHER PROTEIN
 CC PRECURSORS AT SITES COMPRISED OF PAIRS OF BASIC AMINO ACID
 CC RESIDUES. SUBSTRATES INCLUDE POMC, RENIN, ENKEPHALIN, DYNORPHIN,
 CC SOMATOSTATIN AND INSULIN.
 CC -1- CATALYTIC ACTIVITY: Release of protein hormones, neuropeptides and
 CC renin from their precursors, generally by cleavage of -Lys-Arg-|-
 CC bonds.
 CC -1- COFACTOR: CALCIUM DEPENDENT.
 CC -1- SUBCELLULAR LOCATION: LOCALIZED IN THE SECRETION GRANULES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8. FURIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M76705; AAA40945.1; -
 DR EMBL: M83745; AAA41476.1; -
 DR PIR: A41556; KXRTCL.
 DR HSP: Q45670; IDBI.
 DR MEROPS: S08.072; -
 DR InterPro: IPR002884; P_domain.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PR00082; Peptidase_S8; 1.
 DR Pfam: PF01483; P; PARTIAL.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PRODOM: PD000717; P_domain; 1.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KM Hydrolyase; Serine protease; Glycoprotein; Zymogen; Calcium; Signal.
 FT SIGNAL 1 27 POTENTIAL.
 FT PROPEP 28 110 POTENTIAL.
 FT CHAIN 111 752 NEUROENDOCRINE CONVERTASE 1.
 FT DOMAIN 122 410 CATALYTIC.
 FT DOMAIN 739 751 AMPHIPATHIC (POTENTIAL).
 FT ACT_SITE 167 167 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 208 208 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 382 382 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 366 366 T -> Y (IN REF. 2).
 FT CONFLICT 514 514 E -> A (IN REF. 2).
 SQ SEQUENCE 752 AA; 84120 MW; F630AD830A076DED CRC64;
 Query Match 6.7%; Score 85.5; DB 1; Length 752;
 Best Local Similarity 25.1%; Pred. No. 11; Indels 71; Gaps 13;
 Matches 58; Conservative 25; Mismatches 77;
 QY 17 ELDDVNGHK-----FSVSGEGDVTYGLTKLFICTGKLP-----VPMPLVTTFSYG 67
 Db 540 ERTSTNGFRNMFNMGVHTWGENPV--GTMTLKVTDMSGMQEGRIIVMKKIL-----HG 593
 QY 68 VOCFSRPPDHAKRHDFFKSAPEGVQORTIFPKDGNKTRAEVKEFGDTLVNRELK 127
 Db 594 T---SSQPEHMKO-----PRVYTSYNTV-----QNDRGV---EKMNVAVEEKP 631
 QY 128 IDEKFGNLIIGHLEYNYNSHNVYIMADKOKNGIKVNFKIRHNIIEGSOVLADYOOONP 187
 Db 632 TONSLNGNLVLPK---NSSSSVEDRRDROVQAPSKAMLR-----LLOSASFKNTP 680

OY 188 ILDGPVLLPDNHYLSTQSALSKDPNEKRDHVLLEFYTAGITHGMDLYK 238
Db 681 -----SKOS--SKIPSAKLS-----VPYEGLYPALEKLNK 708

RESULT 13
ITR3_MESAU
ID ITR3_MESAU STANDARD; PRT; 886 AA.
AC P97280;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Inter-alpha-trypsin inhibitor heavy chain H3 precursor (ITI heavy chain H3) (Hc3).
GN ITH3.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97420688; PubMed=9276673;
RA Nakatani T., Suzuki Y., Yamamoto T., Sinochana H.;
RT "Molecular cloning and sequencing of cDNAs encoding three heavy-chain precursors of the inter-alpha-trypsin inhibitor in Syrian hamster: implications for the evolution of the inter-alpha-trypsin inhibitor heavy chain family.";
RT J. Biochem. 122:71-82(1997).
RL [2]
RN SEQUENCE OF 31-50; 446-472 AND 504-523, AND SUBUNITS.
RC TISSUE=Plasma;
RX MEDLINE=97018241; PubMed=8864857;
RA Yamamoto T., Yamamoto K., Sinochana H.;
RT "Inter-alpha-trypsin inhibitor and its related proteins in Syrian hamster urine and plasma.";
RT J. Biochem. 120:145-152(1996).
RL [1]
CC -1- FUNCTION: MAY ACT AS A CARRIER OF HYALURONAN IN SERUM OR AS A BINDING PROTEIN BETWEEN HYALURONAN AND OTHER MATRIX PROTEIN, INCLUDING THOSE ON CELL SURFACES IN TISSUES TO REGULATE THE LOCALIZATION, SYNTHESIS AND DEGRADATION OF HYALURONAN WHICH ARE ESSENTIAL TO CELLS UNDERGOING BIOLOGICAL PROCESSES (BY SIMILARITY).
CC -1- SUBUNIT: I-ALPHA-I PLASMA PROTEASE INHIBITORS ARE ASSEMBLED FROM ONE OR TWO HEAVY CHAINS (H1, H2 OR H3) AND ONE LIGHT CHAIN, BIKUNIN. INTER-ALPHA-INHIBITOR (I-ALPHA-I) IS COMPOSED OF H1, H2 AND BIKUNIN. INTER-ALPHA-LIKE INHIBITOR (I-ALPHA-LI) OF H2 AND BIKUNIN, AND PRE-ALPHA-INHIBITOR (P-ALPHA-I) OF H3 AND BIKUNIN.
CC -1- PM: HEAVY CHAINS ARE INTERLINKED WITH BIKUNIN VIA A CHONDROITIN 4-SULFATE BRIDGE TO THEIR C-TERMINAL ASPARTATE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ITH FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D89287; BAA13940.1; -
DR InterPro; IPR002035; VWFA_A.
DR Pfam; PF00092; vwa; 1.
DR SMART; SMO0327; VWFA; 1.
DR PROSITE; PS50234; VWFA; 1.
KW Serine protease inhibitor; Repeat; Signal; Multigene family;
KW Glycoprotein.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 30 BY SIMILARITY.

FT CHAIN 31 646 INTER-ALPHA-TRYPsin INHIBITOR HEAVY CHAIN
FT H3.
FT PROPEP 647 886 BY SIMILARITY.
FT DOMAIN 279 439 VWFA.
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 577 577 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT BINDING 646 646 CHONDROITIN 4-SULFATE, CROSS-LINK SITE (BY SIMILARITY).
FT SO SEQUENCE 886 AA; 99018 MW; AC0594C685257688 CXC64;
Query Match 6.7%; Score 85.5; DB 1; Length 886;
Best Local Similarity 23.8%; Pred. No. 13;
Matches 36; Conservative 34; Mismatches 62; Indels 19; Gaps 7;

OY 74 YPDHAKRDPFKSMPEGYVQRT-----TFKDDGNYKRAEYKFGD---TLVRIE 124
Db 476 YPENAIL-DLTRKNSYPHEDGSEFAVAGRLADSDMNNFK--ADVKGALNDLFTPEVD 532

OY 125 LKSID--FKEDGNLGHKLEYNSHNYIMADKONKIKYFKIRHNIEDGSVQLADY 182
Db 533 MKEDDAKKEGYIFGNTIERMAYLTTEQLERKN--AHGEKENLTAQLLESLKY 589

OY 183 QQNTPIIDGPVLLPDNHYLSTQSALSKDPNE 213
Db 590 HFVPLTPMVTYKPEDN--EDQTSIDAPGE 618

RESULT 14
ITR3B_CHLFR
ID ITR3B_CHLFR STANDARD; PRT; 312 AA.
AC 084101;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thioedoxin reductase (EC 1.6.4.5) (TRXR).
GN TRXR OR CT099.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/Cx;
RX MEDLINE=99008089; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusev R.L., Zhao Q., Koonin E.V., Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.";
RT Science 282:754-759(1998).
CC -1- CATALYTIC ACTIVITY: NADPH + oxidized thioedoxin = NADP(+) + reduced thioedoxin.
CC -1- COFACTOR: FAD.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE OXIDOREDUCTASES CLASS-II.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001284; AAC67690.1; ALT_INIT.
DR HSSP; Q39243; 1YDC.
DR PHCI-2DPAGE; 084101; -
DR InterPro; IPR00759; Adrndx_reductase.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR001100; Pyr_redox.
DR InterPro; IPR000103; Pyridine_redox_2;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 17:54:24 ; Search time 40 Seconds
(without alignments)
572.000 Million cell updates/sec

Title: US-10-071-976-2

Perfect score: 1270

Sequence: 1 MSKGEELFTAVVPILVEIDG.....VLEFVTAAGITHGMDLYK 238

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR_73:*
2: pir1:*
3: pir2:*
4: pir3:*
5: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	97.3	238	1 J01514	green-fluorescent
2	105.5	8.3	785	2 H72228	hypothetical prote
3	100.5	7.9	887	2 E82590	leucyl-tRNA synth
4	99	7.8	336	2 C64468	hypothetical prote
5	92.5	7.3	2573	2 D71614	hypothetical prote
6	90	7.1	471	2 T27856	hypothetical prote
7	90	7.1	797	2 JC4078	protective surface
8	90	7.1	808	2 F64102	protective surface
9	89.5	7.0	393	2 C64613	conserved hypotet
10	88.5	7.0	312	2 C81710	thioredoxin reduct
11	88.5	7.0	425	2 C97354	hypothetical prote
12	87.5	6.9	861	2 H64102	leucine-tRNA ligas
13	87	6.9	822	2 C71633	ATP-dependent nucl
14	87	6.9	2334	2 S32920	cell wall-associat
15	86	6.8	357	2 G81355	tRNA (uracil-5-)-m
16	85.5	6.7	752	2 KKR7C1	proprotein convert
17	85.5	6.7	836	1 J0VJL1	DNA-directed DNA p
18	85.5	6.7	889	2 J05576	inter-alpha-trypsi
19	84.5	6.7	351	2 B71556	probable thioredox
20	84.5	6.7	860	2 AC0582	leucyl-tRNA synth
21	83	6.5	21	2 AD2052	hypothetical prote
22	83	6.5	533	2 S47271	ATMNAK-related prot
23	83	6.5	1092	2 S42798	fibronectin-bindin
24	83	6.5	3472	2 T31308	hypothetical 367k
25	82.5	6.5	613	2 A99552	oligodeoxyribosidase
26	82.5	6.5	1134	2 A60234	Iga Fc receptor pr
27	82.5	6.5	1164	1 FCSOAG	Iga Fc receptor pr
28	82	6.5	207	2 C70483	hypothetical prote
29	82	6.5	653	1 A11535	fructose-Diphosph

30	82	6.5	764	2 T25012	hypothetical prote
31	82	6.5	1082	2 T41988	hypothetical prote
32	82	6.5	1433	2 B83952	DNA polymerase III
33	82	6.5	2166	2 G70163	hypothetical prote
34	81	6.4	647	2 H89988	hypothetical prote
35	81	6.4	953	2 S5156	probable membrane
36	80.5	6.3	355	1 LKCH	ubiquinol-cytochr
37	80.5	6.3	504	1 S23558	hypothetical prote
38	80.5	6.3	636	2 C70031	hypothetical prote
39	80.5	6.3	655	2 D83917	DNA topoisomerase
40	80.5	6.3	1334	2 AB1775	hypothetical prote
41	80.5	6.3	1346	2 G71613	hypothetical prote
42	80.5	6.3	2222	2 A36028	DNA-directed DNA p
43	80.5	6.3	6359	2 T31679	bactracin synthet
44	80	6.3	311	2 C97349	uncharacterized pr
45	80	6.3	393	2 E82897	hypothetical prote

ALIGNMENTS

RESULT 1
J01514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C:Date: 03-Dec-1999 #sequence-revision 03-Dec-1999 #text-change 23-Mar-2001
C:Accession: J050692; J01514; P00335; S48693; S51330; S51331
R:Prascher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.
Gene 111, 229-233, 1992
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A:Reference number: J01514; MUID:92175527; PMID:1347277
A:Accession: J050692
A:Molecule type: DNA
A:Residues: 1-107,'S',109-238 <PRA1>
A:Cross-references: GB:M62654; NID:9155662; PIDN:AAA27722.1; PID:9155663
A:Accession: J01514
A:Molecule type: mRNA
A:Residues: 1-99,'F',101-140,'L',142-218,'V',220-238 <PRA2>
A:Cross-references: GB:M62653; NID:9155660; PIDN:AAA27721.1; PID:9155661
A:Accession: P00335
A:Molecule type: Protein
A:Residues: 46-64;74-122;132-151;154-183;185-200 <PRA3>
R:Inouye, S.; Tsuji, F.I.
FBS Lett. 351, 211-214, 1994
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A:Reference number: S48693; MUID:94364470; PMID:8082767
A:Accession: S48693
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24,'Q',26-156,'P',158-171,'K',173-238 <INO>
A:Cross-references: GB:L29345; NID:9606383; PIDN:AAA58246.1; PID:9606384
R:Watkins, J.N.; Campbell, A.K.
Submitted to the EMBL Data Library, January 1995
A:Reference number: S51330
A:Accession: S51330
A:Molecule type: mRNA
A:Residues: 1-13,'V',15-24,'Q',26-44,'N',46-153,'G',155-156,'P',158-171,'K',173-227,
A:Cross-references: EMBL:X83959; NID:9634008; PIDN:CAA58789.1; PID:9634009
A:Experimental source: clone gtfp
A:Accession: S51331
A:Molecule type: mRNA
A:Residues: 1-24,'Q',26-29,'R',31-83,'L',85-153,'G',155-156,'P',158-171,'K',173-208,
A:Cross-references: EMBL:X83960; NID:9634010; PIDN:CAA58790.1; PID:9634011
A:Experimental source: clone gfp2
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A65692; PDB:1GFL
A:Contents: annotation: X-ray crystallography, 1.9 angstroms, residues 'A',2-79,'R',8
A:Note: engineered sequence based on J01514, cloned and expressed in Escherichia coli
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Nat. Biotechnol. 14, 1246-1251, 1996
A:Title: The molecular structure of green fluorescent protein.
A:Reference number: A58953; MUID:98294543; PMID:9631087

A:Contents: annotation; X-ray crystallography, 1.9 angstroms
 C:Comment: This protein is excited by the photoprotein aequorin (see PIR:AQJFNV) emitting
 C:Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-
 C:Genetics:
 A:Gene: GFP
 A:Introns: 69/3; 167/3
 C:Superfamily: green-fluorescent protein
 C:Keywords: chromoprotein; luminescence
 F:65-67/Cross-link: 5-imidazolium (Ser-Gly) #status experimental
 F:66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 97.3%; Score 1236; DB 1; Length 238;
 Best Local Similarity 96.2%; Pred. No. 5.2e-96;
 Matches 229; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSKGEELTAVVPLIVELDGVNGHKSFSVSGEGSDYTGKLTFFICTTGKLPVPTL 60
 DB 1 MSKGEELTAVVPLIVELDGVNGHKSFSVSGEGSDYTGKLTFFICTTGKLPVPTL 60
 OY 61 VTTFSSVQCFSRYPDHKKRHFESKAMPEGVQORTIFKDDGNKTRAEVKEGDTLV 120
 DB 61 VTTFSSVQCFSRYPDHKKRHFESKAMPEGVQORTIFKDDGNKTRAEVKEGDTLV 120
 OY 121 NRTELKIDREKDNIIHGKLEKYNHNYIMADKKNKYNKTRHNEDGSVOLAD 180
 DB 121 NRTELKIDREKDNIIHGKLEKYNHNYIMADKKNKYNKTRHNEDGSVOLAD 180
 OY 181 YVOQNTFILDGPVLLPNNHLYSTOSALSKDPNEKRDMHVLFEYTAGITGHMDLYK 238
 DB 181 YVOQNTFILDGPVLLPNNHLYSTOSALSKDPNEKRDMHVLFEYTAGITGHMDLYK 238

RESULT 2

H72228
 Hypoetical protein TM1624 - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: H72228
 R:Neelson, K.E.; Clayton, R.A.; Gill, S.R.; Dodson, R.J.; Haft, D.H.; Hickey
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 C.M.
 Nature 399, 333-329, 1999
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
 A:Reference number: A72200; MUID:99287316; PMID:10360571
 A:Accession: H72228
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-785 <ARN>
 A:Cross-references: GB:AE001806; GB:AE00512; NID:94982196; PIDN:AD36691.1; PID:9498219
 A:Experimental source: strain MSB8
 C:Genetics:
 A:Gene: TM1624

Query Match 8.3%; Score 105.5; DB 2; Length 785;
 Best Local Similarity 23.9%; Pred. No. 0.81;
 Matches 43; Conservative 25; Mismatches 65; Indels 47; Gaps 7;

OY 21 DVNCH-----KFSVSGEGSDYTGKLTFFICTTGKLPVPTLVTFSSVQCF 72
 DB 5 DLNGFWSVRDNEGFSEGTVPV-VQADLVK-----GLLPHYVGM----- 46
 OY 73 RYDPMKRHDFESKAMPEGVQORTIFKDDGNKTRAEVKEGDTLVNRTELKIDFKE 132
 DB 47 -----NEDLFKIEDREWIYEREFEFKEDEKGEHVDLFEVDLSDVLLNGVLL- 97
 OY 133 DGNILGHKLKLEYNHNYIMADKKNKYNKTRHNEDGSVOLADYVOQNTFILDGP 192
 DB 98 -GSTEDMKFTEIRFQVNVL-----KKNHLKAVIK-----SPIRVEKTELEQNGVVGCP 145

RESULT 3

E82590
 leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: E82590
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: E82590
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-887 <STM>
 A:Cross-references: GB:AE004031; GB:AE003849; NID:99107309; PIDN:AAF84975.1; GSPDB:GN
 R:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
 as-Neto, E.; Docena, C.; El-Dorri, H.; Fachinani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitejima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
 chado, M.A.; Madelra, A.M.B.N.; Madelra, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri,
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 M.; Tsunako, M.H.; Vailida, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF2176
 C:Superfamily: leucine-tRNA ligase

Query Match 7.9%; Score 100.5; DB 2; Length 887;
 Best Local Similarity 23.2%; Pred. No. 2.5;
 Matches 46; Conservative 28; Mismatches 67; Indels 57; Gaps 10;

OY 49 TTGKLPVPTLVTFSSVQCFSRYPDHKKRHFESKAMPEGVQORTIFKDDGNV-- 106
 DB 329 TNEOLPV-WVANFVLYMGYGAWVPGHQRDEF--ANKYGLPIRQVIALKEPKNDE 385
 OY 107 -----KTRAEVKEGDTLVNRTELKIDFKEKDNIIHGKLEKYNHNYIMADKKN 152
 DB 386 STEPEPVQRMVYADKTR---EFE---LINAEPFGDLQVAFVLAERF----- 429
 OY 153 MADKKNK-IVNFKIRHNIEDGSVOLADYVOQNTFILDGPVPTLVN-----DGPVLLPDN 198
 DB 430 ---RQGRGQRVNVYRLR---DMGVSQRQWGCIPVYICPTGCAVVPEDQVPIILPEN 482
 OY 199 -HYISTOSALSKDPNEKR 215
 DB 483 VAFSGTSPKTPDEWKR 500

RESULT 4

C64468
 Hypoetical protein MJ1348 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: C64468
 R:Bolt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
 ; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klek, H.P.; Frazer, C.M.; Smith, H.O.; Woese
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
 A:Reference number: A64300; MUID:96337999; PMID:8688087
 A:Accession: C64468
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-336 <BDL>
 A:Cross-references: GB:U67574; GB:L77117; NID:91591976; PIDN:AAB99360.1; PID:91591990
 C:Genetics:
 A:Map position: FOR1295121-1296131

A:Start codon: TTG

7.8%; Score 99; DB 2; Length 336;

Query Match
Best Local Similarity 24.3%; Pred. No. 0.93;
Matches 67; Conservative 40; Mismatches 65; Indels 104; Gaps 19;

OY 19 DGDVNGHKSFSVGE-----GEGDVYGGKLTLLFICTYCKLPYV-----57
DB 91 DGDV-----YNSGELSTIASIFAKIGKLDI-----TKNPLMSGGE-----WIINDYKIDAN 137
OY 58 -----PLVYTFST--GVQCFSRYPDMKR-----HDFKSMAPESGYOQRTIF 99
DB 138 SEDILKSVLTFDFSEERKEILNREP--HLRKLFEEDNIYFNNSDFEDFMK-----MF 187
OY 100 FKDDGNYKTRADEVKE-----GDTLVNRIELKGDIPKEDGNIL--GHKLEYNYS 147
DB 188 FIGAGNKKRFLLEVVEEFENKIKSCQISNEYNEI--TKRPD--KMSDLAIATLHLENEYEK 245
OY 148 HNVYIMADK-----ORNGIKVNFKIRHNIEDGSV-----OLADYYOQNTPI 189
DB 246 CLYYVMEKEFEDEDEFNEIKNKINLIVDAIYNKNGNVKKEEWLNLENEYKE---II 302
OY 190 DGPVILPDNHYLSTOSALSAD--PNEKRDMHVLLEFV 224
DB 303 KRP--LPNTY-----KDAHNDLLEILDYVYLKEFI 332

RESULT 5

D71614 hypothetical protein PFB0460c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: D71614

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Perle, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
Science 282, 1126-1132, 1998
A>Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: D71614
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2573 <GAR>
A:Cross-references: GB:AE001396; GB:AE001362; NID:93845188; PIDN:AACT1881.1; PID:9384519
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0460c

Query Match 7.3%; Score 92.5; DB 2; Length 2573;

Best Local Similarity 26.2%; Pred. No. 47;

Matches 34; Conservative 31; Mismatches 52; Indels 13; Gaps 5;

OY 93 VOQRTIFPKD--DGNYKTRAFFEGDILVNRIELKGDIFKEDGNILGHKLEYN--YNSH 148
DB 126 LKKEITLCKDIKSGSNDPMDETSIFKDDMVDDKEIK--DFESSIKIKNKEVYNYIYK 183
OY 149 NVYIMADKONGIKVNFKIRHNIEDGSYOLADYYOQNTPILDGPVILPDNHYLSTOSALS 208
DB 184 NLIHTEKKNKDEKKEKKNNIHNNDENNNM---IYYKN---IDKTHYIIDNNVYHILINDIN 236
OY 209 KDPNEKRDMH 218
DB 237 TYLKREDDYM 246

RESULT 6

T27856 hypothetical protein ZK418.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T27856
R:Fulton, L.

submitted to the EMBL Data Library, April 1994
A:Description: The sequence of C. elegans cosmid ZK418.

A:Reference number: Z20430

A:Accession: T27856

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-471 <FUL>

A:Cross-references: EMBL:U00047; PIDN:AAA50666.1; CESP:ZK418.2

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:ZK418.2

A:Introns: 41/3; 59/1; 139/1; 214/3; 241/2; 295/3; 329/3; 381/3; 454/2

C:Superfamily: Caenorhabditis elegans hypothetical protein ZK418.2

Query Match

7.1%; Score 90; DB 2; Length 471;

Best Local Similarity 23.3%; Pred. No. 8.2;

Matches 49; Conservative 34; Mismatches 81; Indels 46; Gaps 10;

OY 46 FICTGKLPVPPVPLVTFSTFGVQCFSRYPDMKRHDF--FKSAMPE---GYOQRTI 98
DB 140 FLCKFSVLPV-----SHTY-----RYPLEMKKGVPVTFKPEPELINDYKFTVNNEM 186
OY 99 FE---KDD--GNKTRAEVKE-----GDTLVNRIELKGDIFKEDGNILGHKLEYN 146
DB 187 FFDLIKIDNGSYSCDTLVTLDSFKTMTNGVAIINVGMTIGEFYEGYVNLNLEKHAHV 246
OY 147 SHNVYIMADKONGIKV--NFKIRHNIEDGSYOLADYYOQNTPILDGPVIL-----195
DB 247 VPNSKYIKDKIFTEGKVPQCDMWTFYVKNQSP--SDPHVHGVLADSDIVMAAAYVSHMSI 304
OY 196 -PDNHYLSTOSALSADPNEKRDMHVLLEFV 224
DB 305 GPQIAMISHENRLLFNLKSRDMHVLTYI 334

RESULT 7

JC4078

protective surface antigen D-15 precursor - Haemophilus influenzae (type b)

C:Species: Haemophilus influenzae

A:Variety: type b

C:Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999

C:Accession: JC4078

R:Flack, F.S.; Loosmore, S.; Chong, P.; Thomas, W.R.

Gene 156, 97-99, 1995

A>Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus 1

A:Reference number: JC4078; MUID:95255676; PMID:7737523

A:Accession: JC4078

A:Molecule type: DNA

A:Residues: 1-797 <FLA>

A:Cross-references: GB:U13961; NID:9537447; PIDN:AAA85645.1; PID:9537448

A:Experimental source: type b

C:Superfamily: protective surface antigen D-15

C:Keywords: surface antigen

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match

7.1%; Score 90; DB 2; Length 797;

Best Local Similarity 20.2%; Pred. No. 16;

Matches 52; Conservative 36; Mismatches 96; Indels 74; Gaps 12;

OY 22 VNGHK-----FVSGEGBDVYTKLILKFICTTGKLPVPPVLTTFSTFGVQCFSRYP 75
DB 402 INGSNDEVDVYKVKERTGSIINF-----IGYTESGISYO 438
OY 76 DHMKRHDF-----FKSAMPEGYVOQ-----RTIFFKDDGNYKTRA 110
DB 439 ASVKODNPLGCGAASVIGTKNDYGTSVNLGTEPYFTKDGYSLOGNFFEFYNDKSDYF 498
OY 111 EYKFGDPLVNRIELKGDIFKEDGNI---LGH-----KLEYNNSHNVYIMADKOK- 158
DB 499 SSNRYKRTYGSNVL--GPPVNNNSYVYGLGTYTKISFALEYN---RNLYIQSMKFKG 554
OY 159 NGIKVY--FKIRHNIEDGSYOLADYYOQNTPI--LDGPVILP--DNHYLSTOSALSADPNEK 214
DB 555 NGIKTNDPFSFGWYNSLNRGYPPTKGVKASLAGRVITPGSDNRYKLSADVQGFYPLD 614

OY 215 RDHMLLEFVTAAGITG 232
 Db 615 RDHMLVVSASAKSAGYANG 632

RESULT 8

F64102
 C:Species: Haemophilus influenzae
 C:Date: 13-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1998
 C:Accession: F64102
 R:Flischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M.; Science 269, 496-512, 1995
 A:Authors: Guelm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MUID:95350630; PMID:7542800
 A:Accession: F64102
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-808 <TIGR>
 A:Cross-references: GB:AE000587; TIGR:HT0917
 C:Superfamily: protective surface antigen D-15
 C:Keywords: surface antigen

Query Match

Best Local Similarity 7.1%; Score 90; DB 2; Length 808;
 Matches 52; Conservative 36; Mismatches 96; Indels 74; Gaps 12;

OY 22 VNGHK-----FVSQSGEDDYVYKLTFLCTTGKLPVWPVLTFTSYGOCESRRYP 75
 Db 415 INGSNDVDVYVYKERNKTSING-----IGTSGSISYQ 451
 OY 76 DHMKRHF-----FKSAPGEGYVQ-----RTIFFDQNTYTRA 110
 Db 452 ASYKQDNFLGTGAANVSTAGTKNDYGTVMNGITPTTKQCVSLGAVFENYDMSKSDT 511
 OY 111 EVFEGDVLNRIELKIDFKEDGNI---LGH-----KLEYNASHNYIMADKOK- 158
 Db 512 SSNFKRTTYSNVTL-GFPVNNNSYVGLGHTYKNSPALERN---RNLVIGSMKFKG 567
 OY 159 NGIKVN-FKIRHNEDSSVOLADYYQNTPI-LDGPLYLP--DNHYISTOSALSKDNEK 214
 Db 568 NGIKTNDGDFSGMNNYSLKRGYFPTKGVKASLGGRTVITGSDMKYKLSADYGFYPLD 627
 OY 215 RDHMLLEFVTAAGITG 232
 Db 628 RDHMLVVSASAKSAGYANG 645

RESULT 9

C64613
 C:Species: Helicobacter pylori
 C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
 C:Accession: C64613
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Flischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenna, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Matthey, L.; Nature 388, 539-547, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.; Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A:Reference number: A64520; MUID:97394467; PMID:9252185
 A:Accession: C64613
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-393 <TOM>
 A:Cross-references: GB:AE000587; GB:AE000511; NID:92313869; PIDN:AAD07796.1; PID:9231387

Query Match 7.0%; Score 89.5; DB 2; Length 393;
 Best Local Similarity 20.8%; Pred. No. 7.1;

Matches 51; Conservative 42; Mismatches 77; Indels 75; Gaps 15;

OY 34 EGDVYTKLTLK-----FICTGKLPVWPVLTFTSYGOCESRRYPDKRH-- 81
 Db 174 QGD---GRVLSESPNHRCEKIFV---HPPVPM-----NEKKHR 207
 OY 82 ----DFKSAW----PGYVOQRT---IFEKD-----GNKYRAEVKESD-TLVRIE 124
 Db 208 VLSEKFLNEALRYLKPFGFELRFDSDLYFEDSLKALKNFQCEIEFKKNAQIPVSKYE 267
 OY 125 LKSIDFKED-GNIGHKLEYN---YNSHNYIMADK---OKNGIKVNFKIRHNEDG-SV 176
 Db 268 ARMKKLKDDYDLKITYLSLENNETPFNDHADFSDFTTTSKSVGTLTKTKIIOEGYFV 327
 OY 177 QLADYQOQNTPI-----DGP-----VLLPDNHYLSTOSALSKDPEKRDHMLLEFV 224
 Db 328 HVCATYENKQDFVELSMGDFDWRVRLFVLTENQIYLNKSPILKTLNNKHAHLLONIL 387
 OY 225 TAAGI 229
 Db 388 SOKGI 392

RESULT 10

C81710
 C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
 C:Accession: C81710
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gail, S.R.; Heidelberg, J.F.; White, O.; Hicke, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salze, Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
 A:Reference number: A81500; MUID:20150255; PMID:10684935
 A:Accession: C81710
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-312 <TET>
 A:Cross-references: GB:AE002304; GB:AE002160; NID:97190409; PIDN:AAF39233.1; PID:9719
 A:Genetic source: strain Nigg (MOPn)
 A:Gene: TC0375
 C:Superfamily: thioredoxin reductase; thioredoxin reductase homology

Query Match

Best Local Similarity 7.0%; Score 88.5; DB 2; Length 312;
 Matches 48; Conservative 29; Mismatches 81; Indels 69; Gaps 9;

OY 1 MSKGEELFNAVPLVLDDGVNGHKFSYSGEGSDVYTKLTKFICTTGKLPVWPPTL 60
 Db 97 LKSGEDFTCDACIAT--GASAKRLISIPGADNFERMOKGVYCAVC-DGASPI----- 147
 OY 61 VTFESYGVQCESRYPDMKRRHDF-----KSAPGEGYVQRTIFFKDGNVYKTRAEVKE 115
 Db 148 -----FRDRDLFVIGGGSALAEAMFLTR-----YGRVYVHR 161
 OY 116 GDTL-----VNRLEKIDFKEDGNTLGHLEYNASHNYIMADKQNGIKVNEKIRH 169
 Db 182 RDTLRASKAMVWNAQ-----ANKKIYFIAMSEVVKILGSLVRSI-----DIFN 225
 OY 170 NIEDGSVQL-----ADYQOQNTPIIDGVVLLPDNHYLSTOSALS 209
 Db 226 NVEKTYTMEAGVFFAIGHQPTAFVIGSGLSIDENGYITTEKSSSR 272

RESULT 11

C97354
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: C97354
 R:Nolling, U.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, O.; Gibson, R.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C1C
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: C97354
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-425 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK61622.1; PID:g15026806; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC3702

Query Match 7.0%; Score 88.5; DB 2; Length 425;
 Best Local Similarity 24.3%; Pred. No. 9.6;
 Matches 59; Conservative 29; Mismatches 82; Indels 73; Gaps 14;

QY 20 GDVNGHKEFSVSGEGEDVYTKLTKFCITGK-----LP-----VPM 57
 DB 109 GFVNGKLLPASGEAIKD--FGKV---LKNKGKDLKGLTRQVEVEMFGVGRVSGPNIIEF 162
 QY 58 PLVTTFSTYGVQCSRRPDHMKRHFESKAMPEGYVOQRITFFDDGNYKTR-----AEV 112
 DB 163 SEVVKRNKNGVDI-----KDNFVAKVSEDKVGNP---SGEGYSTEEWYNYLKE 211
 QY 113 KFEED-----TLVNRLELKGIDPKE--DGNILGHKLEYNY-----NSH- 148
 DB 212 KYGDANYLLCTDEKSLNSSEKIDSFRRKNGVYG-KTAVAYGENLDIVSAMTSNSHO 270
 QY 149 -NYIMADKOKNGIKVFKIRHNIEDSGVOLADYVOQNTPIIDGVPVLLPDNHLYSTOSAL 207
 DB 271 YNPILTDGEKFDKNNMYNKLKEYGDESVAYSKYV-EHTLTEDYIKTLPDEYF--TSKGL 327
 QY 208 SKD 210
 DB 328 VKD 330

RESULT 12

H64102
 leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)
 N:Alternate names: leucyl-tRNA synthetase
 C:Species: Haemophilus influenzae
 C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 03-Jun-2002
 C:Accession: H64102
 R:Rieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Meldrum, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, N.S.M.
 Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MUID:95350630; PMID:7542800
 A:Accession: H64102
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-861 <TIGR>
 A:Cross-references: GB:U32774; GB:U42023; NID:g1573942; PIDN:AAK22581.1; PID:g1573943; T
 C:Genetics:
 A:Superfamily: leucine-tRNA ligase
 C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 6.9%; Score 87.5; DB 2; Length 861;
 Best Local Similarity 24.3%; Pred. No. 29;
 Matches 45; Conservative 25; Mismatches 68; Indels 47; Gaps 9;

QY 49 TTGKLVPWPVLTTFSTYGVQCSRRPDHMKRHFESKAMPEGYVOQRITFFKDDGKNT 108
 DB 314 TGDRLPI-WVANFYLIMHYGTGAVAVPAHQDQD--FEPAQYSLPIKQVIAPLADDEIDL 370
 QY 109 RAEVKFEGLTVNRIELKIGDKEDGNLGHKLEYNNSHNVYIMADK-OANGT---KVN 164
 DB 371 TKQAFVEHGKLVNSDEDEKGNF--DGAENG-----TADLLEKLGKGRQVN 414

QY 165 FKIRH-----NIEDGSVOLADYVOQNTPIIDGVPVLLPDNHLY-STOSAL 207
 DB 415 YRLDKMGVSKQRTWGAFTPLTLFENGDVFA-----PHEDDLPIILPEVDYMDGVASPI 467
 QY 208 SKDPN 212
 DB 468 NADPN 472

RESULT 13

C71633
 ATP-dependent nuclease chain A (addA) RP734 - Rickettsia prowazekii
 C:Species: Rickettsia prowazekii
 C>Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
 C:Accession: C71633
 R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Slicheritz-Ponten, T.; Alsmark
 Nature 396, 133-140, 1998
 A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
 A:Reference number: A71630; MUID:99039499; PMID:9823893
 A:Accession: C71633
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-822 <AND>
 A:Cross-references: GB:AJ235273; GB:AJ235269; NID:g3861237; PIDN:CAA15163.1; PID:e134
 A:Experimental source: strain Madrid E
 C:Genetics:
 A:Gene: addA; RP734

Query Match 6.9%; Score 87; DB 2; Length 822;
 Best Local Similarity 25.2%; Pred. No. 30;
 Matches 54; Conservative 28; Mismatches 54; Indels 78; Gaps 13;

QY 57 WPLVTTFSTYGVQCSRRPDHMKRHFESKAMPEGYVOQRITFFKDDGNTYRAEVFEG 116
 DB 473 WPLVT-----NOKQKEFFWT-LPE-----DKNMAKADLTL-- 503
 QY 117 DTLVN-----RIELKGI-----DFKEDGNILGHKLEYNNSHNVYIMADKOKNGIKVNF 165
 DB 504 DKIVNFIKEKIKSGILSTSRSEKDFMLVAKRD--KFSHN--LIKELSKAKLVEI 559
 QY 166 KIRHNIEDGSVOLADYVOQNTPIID-----GPVLLPDNHLYSTOSALSKDP----- 211
 DB 560 SDRINL-----KENPIIDLLAAKFAVLLPDDDL--NLACLKSPITIGISQKL 606
 QY 212 -----NEKRDHNVLEFVTAAGITHGMD---ELXK 238
 DB 607 YTLVAKKNDHLMVLSHNDIYHKLDSTIEYK 640

RESULT 14

S32920
 cell wall-associated protein precursor wapa (similarity) - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Jul-2000
 C:Accession: S32920; E69730; T47101
 R:Poster, S.J.
 Mol. Microbiol. 8, 299-310, 1993
 A:Title: Molecular analysis of three major wall-associated proteins of Bacillus subtilis
 A:Reference number: S32919; MUID:93302506; PMID:8316082
 A:Accession: S32920
 A:Molecule type: DNA
 A:Residues: 1-2334 <POS>
 A:Cross-references: GB:L05634; NID:g304117; PIDN:AAA22883.1; PID:g304179
 R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 A:; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A:; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 lecher, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Huijo, M
 koetter, J.; Konigstein, G.; Krogh, K.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot, A.; Kuchel, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A. A.; Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. A:Reference number: A69380; MUID:9804033; PMID:9384377
 A:Accession: E69730
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2334 <RUN>
 A:Cross-references: GB:AL009126; NID:92636442; PIDN:CAB1595.1; PID:92636469
 A:Experimental source: strain 168
 R:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.
 Microbiology 141, 337-343, 1995
 A:Title: Cloning and sequencing of a 29 kb region of the *Bacillus subtilis* genome contai
 A:Reference number: 224350; MUID:95219088; PMID:7704263
 A:Accession: T47101
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2334 <YOS>
 A:Cross-references: EMBL:D31856; NID:9603765; PIDN:BAA06656.1; PID:9603782
 A:Experimental source: strain BGSC1A1
 C:Genetics:
 A:Gene: wapa; N17G
 C:Superfamily: cell wall-associated protein wapa

Query Match 6.9%; Score 87; DB 2; Length 2334;
 Best Local Similarity 19.3%; Pred. No. 1.2e+02;
 Matches 58; Conservative 41; Mismatches 105; Indels 96; Gaps 12;

OY 21 DVNGHRSVSGEGEDVY--GKLTKEICTTGKLPWPMTLVTFESYQV----- 69
 Db 1054 DENGHYVSTGPKKKKTYSENDLKKVDTDS-----YVTSYDSEGRIVKQYS 1105
 OY 70 -----CFSRYPDHKKRHDFFKSAMEGV-----QQRITF-----FKDDG 104
 Db 1106 ANSTKAPVFTETQYSGHLEKALNAKKEITYVSYDADKTLTLPNGRKRVOYGYNEAG 1165
 OY 105 N-----YKTRAEVKFEEDTLVNRIE-----LKSIDFKEDGNILGHR----- 140
 Db 1166 NPQOVIDADAGLKTITNTKYEAGNVVEDPNDVGTGKATESYQYDKDGNVTSKDAVGT 1225
 OY 141 LETYNASHN-VYIMADKKN-----GIVNEKIRHNIEDGSVOLADYV-----QONTPI 188
 Db 1226 ETEYENKNNNDVTYKKDEGVNTDIAYDGLDAVSETDSGKSSSAAYVYDKYQNOIOSKDL 1285
 OY 189 LDGPVLLPDNHYLSTOSALSKDPRER-----DHNVLEFVTAAGITHG 232
 Db 1286 SASITNIIKDSFEAKSGMNLTKSKDRKISVIADKSGVLSGSKALEVLSOSTSAGTDHG 1345

RESULT 15
 G81355
 tRNA (uracil-5-)-methyltransferase (EC 2.1.1.35) Cj0831c [imported] - Campylobacter jej
 C:Species: Campylobacter jejuni
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C:Accession: G81355
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chilling
 C.W.; Ouali, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals h
 A:Reference number: A81250; MUID:20150912; PMID:10688204
 A:Accession: G81355
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-357 <PAR>
 A:Cross-references: GB:AL139076; GB:AL111168; NID:96968128; PIDN:CAB73096.1; PID:9696827
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: trna; Cj0831c
 C:Keywords: methyltransferase; s-adenosylmethionine

Query Match 6.8%; Score 86; DB 2; Length 357;
 Best Local Similarity 21.8%; Pred. No. 12;
 Matches 36; Conservative 23; Mismatches 44; Indels 62; Gaps 6;

OY 79 KRNDFFKSAMEGVQORTTFEFDNDGNYKRAVYK--EGDITV----- 120
 Db 14 EKHSFIKKYKFEYTDKFLFASKDKHYRTRAEISFHEMDTLFAMFDPKSKKYYIEY 73
 OY 121 -----NRIELK--GIDFKEDGNILGKLEYNV----- 146
 Db 74 LDFADEKICAFMRLELYLQDNKKIKKELGVETITTKQELSTLLYHKNIEDIKSNLEN 133
 OY 147 -SHNVYIMADKKNKNGIKVNFKIRH-----NIEDGSVOLADYQON 185
 Db 134 LSNILHINILARSKGKRLIFKTENLRQTLNIDORKI----FYEFN 174

Search completed: July 11, 2003, 17:59:33
 Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 17:43:48 ; Search time 72 Seconds
(without alignments)
440.467 Million cell updates/sec

Title: US-10-071-976-2

Perfect score: 1270

Sequence: 1 MSKGELEFNAVVPILVELDG.....VLEFVTAAGITGMDLELYK 238

Scoring table:

BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.GeneSeq_101002:*

- 1: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*
- 2: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*
- 3: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*
- 4: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*
- 5: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*
- 6: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*
- 7: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*
- 8: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*
- 9: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*
- 10: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*
- 11: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*
- 12: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*
- 13: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*
- 14: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*
- 15: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*
- 16: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*
- 17: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*
- 18: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*
- 19: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*
- 20: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*
- 21: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*
- 22: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*
- 23: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1270	100.0	238	19	AAW52313
2	1267	99.8	238	19	AAW52328
3	1267	99.8	238	19	AAW52330
4	1266	99.7	238	19	AAW52333
5	1266	99.7	238	19	AAW52335
6	1265	99.6	238	19	AAW52351
7	1265	99.6	238	19	AAW52335
8	1265	99.6	238	19	AAW52336
9	1265	99.6	238	19	AAW52337
10	1265	99.6	238	19	AAW52338

11	1265	99.6	238	19	AAW52344	Engineered green f
12	1265	99.6	238	19	AAW52346	Engineered green f
13	1265	99.6	238	19	AAW52349	Engineered green f
14	1265	99.6	238	19	AAW52324	Engineered green f
15	1265	99.6	238	19	AAW52331	Engineered green f
16	1265	99.6	238	19	AAW52334	Engineered green f
17	1265	99.6	238	19	AAW52314	Engineered green f
18	1264	99.5	238	19	AAW52339	Engineered green f
19	1264	99.5	238	19	AAW52340	Engineered green f
20	1264	99.5	238	19	AAW52347	Engineered green f
21	1264	99.5	238	19	AAW52350	Engineered green f
22	1264	99.5	238	19	AAW52332	Engineered green f
23	1263	99.4	238	19	AAW52341	Engineered green f
24	1263	99.4	238	19	AAW52345	Engineered green f
25	1263	99.4	238	19	AAW52348	Engineered green f
26	1262	99.4	238	19	AAW52342	Engineered green f
27	1262	99.4	238	19	AAW52323	Engineered green f
28	1261	99.3	238	19	AAW52343	Engineered green f
29	1261	99.3	238	19	AAW52325	Engineered green f
30	1260	99.2	238	19	AAW52316	Engineered green f
31	1260	99.2	238	19	AAW52317	Engineered green f
32	1257	99.0	238	19	AAW52329	Engineered green f
33	1256	98.9	238	19	AAW52321	Engineered green f
34	1256	98.9	238	19	AAW52322	Engineered green f
35	1253	98.7	238	19	AAW52319	Engineered green f
36	1250	98.4	238	19	AAW52318	Engineered green f
37	1249	98.3	238	19	AAW52320	Engineered green f
38	1246	98.1	238	17	AAW05304	Green fluorescent
39	1246	98.1	238	18	AAW24232	Aequorea victoria
40	1246	98.1	238	19	AAW76371	A. victoria green
41	1246	98.1	238	19	AAW76105	A. victoria green
42	1246	98.1	238	19	AAW40479	Wild-type green fl
43	1246	98.1	238	22	AAE16038	Aequorea victoria
44	1246	98.1	238	19	AAE16038	Aequorea victoria
45	1245	98.0	238	19	AAW65084	A. victoria green

ALIGNMENTS

RESULT 1
AAW52313
ID AAW52313 standard; Protein; 238 AA.
XX
AC AAW52313;
XX
DT 10-JUL-1998 (first entry)
XX
DE Aequorea green fluorescent protein.
XX
KW Green fluorescent protein; engineered fluorescent protein;
KW Reporter molecule; immunological assay; protein-protein interaction;
KW Fluorescence resonance energy transfer system; FRET system.
XX
OS Aequorea victoria.
XX
PN WO9806737-A1.
XX
PD 19-FEB-1998.
XX
PF 15-AUG-1997; 97WO-US14593.
XX
PR 30-AUG-1996; 96US-0706408.
XX
PR 16-AUG-1996; 96US-0024050.
XX
PA (AURO-) AURORA BIOSCIENCES.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PA (UOYOR-) UNIV OREGON.
XX
PI Cubitt AB, Heim R, Ormo MF, Remington JS, Tsien R;
DR WPI; 1998-159454/14.
DR N-PSDB; AAV19946.

XX Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants

PS Claim 1; Page 53; 120pp; English.

XX This sequence is the wild type Aequorea victoria green fluorescent
 CC protein. The nucleic acid was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H, Y,
 CC W or F) and having fluorescent properties different from GFP. (II) are
 CC useful as reporter molecules in immunological or hybridisation assays,
 CC for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.

XX Sequence 238 AA:

SO Query Match 100.0%; Score 1270; DB 19; Length 238;
 Best Local Similarity 100.0%; Pred. No. 2e-123;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVNPILVELDGVNGHKFSVSGEGSDVYTKLTKFICTGKLPVMPPTL 60
 DB 1 MSKGEELFTAVNPILVELDGVNGHKFSVSGEGSDVYTKLTKFICTGKLPVMPPTL 60
 QY VTFPSYGVQCFSRYPDHMKRHDFFKSAPEGYVOQRTIFFKDDGNTKTRAEVFEGDTLV 120
 DB 61 VTFPSYGVQCFSRYPDHMKRHDFFKSAPEGYVOQRTIFFKDDGNTKTRAEVFEGDTLV 120
 QY 121 NRIELKGIDKEDNGNIIGHKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 DB 121 NRIELKGIDKEDNGNIIGHKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 QY 181 YVOQNPPIIDGPVLLPDNHNLTSTQSALSKDPNEKRDMVLEFVTAAGITHGMDLYK 238
 DB 181 YVOQNPPIIDGPVLLPDNHNLTSTQSALSKDPNEKRDMVLEFVTAAGITHGMDLYK 238

RESULT 2
 AAM52328
 ID AAM52328 standard; Protein; 238 AA.
 AC AAM52328;
 XX 10-JUL-1998 (first entry)
 DE Engineered green fluorescent protein S65T.
 XX
 KW Green fluorescent protein; engineered fluorescent protein; mutain;
 KM reporter molecule; immunological assay; protein-protein interaction;
 XX fluorescence resonance energy transfer system; FRET system.
 OS Synthetic.
 OS Aequorea victoria.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 65 /label="S65T"
 FT /note="Ser to Thr mutation"
 XX
 XX WO9806737-A1.
 XX
 XX 19-FEB-1998.

XX 15-AUG-1997; 97WO-US14593.
 XX 30-AUG-1996; 96US-0706408.
 PR 16-AUG-1996; 96US-0024050.
 XX
 XX (AURO-) AURORA BIOSCIENCES.
 PA (REGC) UNIT CALIFORNIA.
 PA (UYOR-) UNIT OREGON.
 PI Cuditt AB, Heim R, Otmio MF, Remington JS, Tsien RY;
 XX WPI; 1998-159454/14.
 XX
 PT Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants

PS Claim 5; Page 7; 120pp; English.

XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52313. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.

XX Sequence 238 AA:

SO Query Match 99.8%; Score 1267; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 4.1e-123;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVNPILVELDGVNGHKFSVSGEGSDVYTKLTKFICTGKLPVMPPTL 60
 DB 1 MSKGEELFTAVNPILVELDGVNGHKFSVSGEGSDVYTKLTKFICTGKLPVMPPTL 60
 QY 61 VTFPSYGVQCFSRYPDHMKRHDFFKSAPEGYVOQRTIFFKDDGNTKTRAEVFEGDTLV 120
 DB 61 VTFPSYGVQCFSRYPDHMKRHDFFKSAPEGYVOQRTIFFKDDGNTKTRAEVFEGDTLV 120
 QY 121 NRIELKGIDKEDNGNIIGHKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 DB 121 NRIELKGIDKEDNGNIIGHKLEYNNSHNHYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 QY 181 YVOQNPPIIDGPVLLPDNHNLTSTQSALSKDPNEKRDMVLEFVTAAGITHGMDLYK 238
 DB 181 YVOQNPPIIDGPVLLPDNHNLTSTQSALSKDPNEKRDMVLEFVTAAGITHGMDLYK 238

RESULT 3
 AAM52330
 ID AAM52330 standard; Protein; 238 AA.
 AC AAM52330;
 XX 10-JUL-1998 (first entry)
 DE Engineered green fluorescent protein S65A.
 XX
 XX Green fluorescent protein; engineered fluorescent protein; mutain;

KM reporter molecule; immunological assay; protein-protein interaction;
 KM fluorescence resonance energy transfer system; FRET system.
 XX Synthetic.
 OS Aequorea victoria.
 XX
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 65 /Label= S65A
 FT /note= "Ser to Ala mutation"
 FT
 XX
 XX W09806737-A1.
 PN
 XX
 XX 19-FEB-1998.
 PD
 XX
 XX 15-AUG-1997; 97WO-US14593.
 PE
 XX
 XX 30-AUG-1996; 96US-0706408.
 PR 16-AUG-1996; 96US-0024050.
 XX
 XX (AURO-) AURORA BIOSCIENCES.
 PA (RECC) UNIV CALIFORNIA.
 PA (UYOR-) UNIV OREGON.
 PI
 XX
 XX Cubitt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
 DR WPI: 1998-159454/14.
 XX
 XX Nucleic acid encoding mutant green fluorescent proteins having
 PF longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PI related vectors and transformants
 XX
 XX Claim 5; Page -: 120pp; English.
 PS
 XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52333. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.
 CC
 XX Sequence 238 AA:
 SQ
 Query Match 99.8%; Score 1267; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 4, 1e-123;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 181 YYOQNTPLIDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITGHMDLYK 238
 RESULT 4
 AAM52333
 ID AAM52333 standard; Protein: 238 AA.
 XX
 XX
 XX AAM52333;
 AC
 XX
 XX 10-JUL-1998 (first entry)
 DT
 XX
 XX Engineered green fluorescent protein Y66F.
 DE
 XX
 XX Green fluorescent protein; engineered fluorescent protein; muten;
 KM reporter molecule; immunological assay; protein-protein interaction;
 KM fluorescence resonance energy transfer system; FRET system.
 XX
 XX Synthetic.
 OS Aequorea victoria.
 XX
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 66 /Label= Y66F
 FT /note= "Tyr to Phe mutation"
 FT
 XX
 XX W09806737-A1.
 PN
 XX
 XX 19-FEB-1998.
 PD
 XX
 XX 15-AUG-1997; 97WO-US14593.
 PE
 XX
 XX 30-AUG-1996; 96US-0706408.
 PR 16-AUG-1996; 96US-0024050.
 XX
 XX (AURO-) AURORA BIOSCIENCES.
 PA (RECC) UNIV CALIFORNIA.
 PA (UYOR-) UNIV OREGON.
 PI
 XX
 XX Cubitt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
 DR WPI: 1998-159454/14.
 XX
 XX Nucleic acid encoding mutant green fluorescent proteins having
 PF longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PI related vectors and transformants
 XX
 XX Claim 5; Page -: 120pp; English.
 PS
 XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52333. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.
 CC
 XX Sequence 238 AA:
 SQ
 Query Match 99.7%; Score 1266; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 3, 3e-123;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;


```

QY 1 MSKGEELFTAVVPIVLVDGVDNGHKFSVSGEGSDVYTGKTLTKICTGKLPVMPPTL 60
DB 1 MSKGEELFTAVVPIVLVDGVDNGHKFSVSGEGSDVYTGKTLTKICTGKLPVMPPTL 60
QY 61 VTTFSYGVOCFSRYPDHKKRHDFFKSAMPEGYVOQRTIFFKDDGNYKTRAEVKEGDTLY 120
DB 61 VTTFSYGVOCFSRYPDHKKRHDFFKSAMPEGYVOQRTIFFKDDGNYKTRAEVKEGDTLY 120
QY 121 NRIELKGIDFEKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFEKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
QY 181 YVQONTPLIDGPVLLPDNHYLSTOSALSKDPNEKRDHVLLEFVTAAGITGMDLYK 238
DB 181 YVQONTPLIDGPVLLPDNHYLSTOSALSKDPNEKRDHVLLEFVTAAGITGMDLYK 238

RESULT 5
AAW52315
ID AAW52315 standard; Protein; 238 AA.
XX
AC AAW52315;
XX
DE 10-JUL-1998 (first entry)
XX
DE Engineered green fluorescent protein.
XX
KW Green fluorescent protein; engineered fluorescent protein; muten;
KW Reporter molecule; immunological assay; protein-protein interaction;
KW fluorescence resonance energy transfer system; FRET system.
XX
OS Synthetic.
OS Aequorea victoria.
XX
XX
XX Key Location/Qualifiers
XX MISC-difference 65
XX /Label- Gly, Thr, Ala, Leu, Cys, Val, Ile
XX
XX W09806737-A1.
XX
XX 19-FEB-1998.
XX
XX 15-AUG-1997; 97WO-US14593.
XX
XX 30-AUG-1996; 96US-0706408.
XX
XX 16-AUG-1996; 96US-0024050.
XX
XX (AURO-) AURORA BIOSCIENCES.
XX (REGC ) UNIV CALIFORNIA.
XX (UYOR-) UNIV OREGON.
XX
XX Cuditt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
XX WPI; 1998-159454/14.
XX
XX
XX Nucleic acid encoding mutant green fluorescent proteins having
XX longer wavelength emission - used as markers for probes and as
XX components of fluorescent resonant energy transfer systems, also
XX related vectors and transformants
XX
XX Claim 2; Page -: 120pp; English.
XX
XX This sequence is an engineered Aequorea victoria green fluorescent
XX protein (GFP) of the invention, created from the sequence shown in
XX AAW52313. The GFP DNA was mutated to produce the nucleic acids (I) of
XX the invention, which encode functional, engineered fluorescent proteins
XX (II) having largely the same sequence as Aequorea green fluorescent
XX protein (GFP) but differing by at least the substitution T203X (X = H,
XX Y, W or F) and having fluorescent properties different from GFP. (II)
XX are useful as reporter molecules in immunological or hybridisation
XX assays, for monitoring proteins in cells and detecting induction of
XX transcription. They are also useful as components of FRET (fluorescence
XX resonance energy transfer) systems, e.g. for detecting protein-protein

```

```

CC interactions, cleavage of substrates and changes in potential across a
CC membrane, and further for making fluorescent substrates for protein
CC kinase. (II) are easily distinguished from known green and blue
CC fluorescent proteins, so allow simultaneous measurements of at least 2
CC proteins in a cell. They emit, and are excited, at relatively long
CC wavelengths where phototoxicity and auto-fluorescence are reduced.
XX
XX Sequence 238 AA:
XX
XX Query Match 99.7%; Score 1266; DB 19; Length 238;
XX Best Local Similarity 99.6%; Pred. No. 5.3e-123;
XX Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 MSKGEELFTAVVPIVLVDGVDNGHKFSVSGEGSDVYTGKTLTKICTGKLPVMPPTL 60
DB 1 MSKGEELFTAVVPIVLVDGVDNGHKFSVSGEGSDVYTGKTLTKICTGKLPVMPPTL 60
QY 61 VTTFSYGVOCFSRYPDHKKRHDFFKSAMPEGYVOQRTIFFKDDGNYKTRAEVKEGDTLY 120
DB 61 VTTFSYGVOCFSRYPDHKKRHDFFKSAMPEGYVOQRTIFFKDDGNYKTRAEVKEGDTLY 120
QY 121 NRIELKGIDFEKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFEKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
QY 181 YVQONTPLIDGPVLLPDNHYLSTOSALSKDPNEKRDHVLLEFVTAAGITGMDLYK 238
DB 181 YVQONTPLIDGPVLLPDNHYLSTOSALSKDPNEKRDHVLLEFVTAAGITGMDLYK 238

RESULT 6
AAW52351
ID AAW52351 standard; Protein; 238 AA.
XX
AC AAW52351;
XX
DE 10-JUL-1998 (first entry)
XX
DE Engineered green fluorescent protein.
XX
KW Green fluorescent protein; engineered fluorescent protein; muten;
KW Reporter molecule; immunological assay; protein-protein interaction;
KW fluorescence resonance energy transfer system; FRET system.
XX
OS Synthetic.
OS Aequorea victoria.
XX
XX
XX Key Location/Qualifiers
XX MISC-difference 224
XX /Label- His, Asn, Gln, Thr, Phe, Trp, Tyr
XX
XX W09806737-A1.
XX
XX 19-FEB-1998.
XX
XX 15-AUG-1997; 97WO-US14593.
XX
XX 30-AUG-1996; 96US-0706408.
XX
XX 16-AUG-1996; 96US-0024050.
XX
XX (AURO-) AURORA BIOSCIENCES.
XX (REGC ) UNIV CALIFORNIA.
XX (UYOR-) UNIV OREGON.
XX
XX Cuditt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
XX WPI; 1998-159454/14.
XX
XX Nucleic acid encoding mutant green fluorescent proteins having
XX longer wavelength emission - used as markers for probes and as
XX components of fluorescent resonant energy transfer systems, also
XX related vectors and transformants

```


PS Claim 47; Page -: 120pp; English.

XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52313. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.

XX Sequence 238 AA;

S0

Query Match 99.6%; Score 1265; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6.7e-123;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 MSKGEELFTAVVPLIVELDGVNGHKFSVSGEGSDVYGGKLTITCTTKLTPVWPTL 60
 Db 1 MSKEELFTAVVPLIVELDGVNGHKFSVSGEGSDVYGGKLTITCTTKLTPVWPTL 60

Oy 61 VTFSYGYQCFSRYPDHKKRHDFFKSAPEGYVOORTIFFDDGNYKTRAVEKFGDPLV 120
 Db 61 VTFSYGYQCFSRYPDHKKRHDFFKSAPEGYVOORTIFFDDGNYKTRAVEKFGDPLV 120

Oy 121 NRIELKGIDFEKEDGNILGHKLEYNNSHNYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
 Db 121 NRIELKGIDFEKEDGNILGHKLEYNNSHNYIMADKONGIKVNFKIRHNIEDGSVOLAD 180

Oy 181 YVOONTPIIDGPVLLPDPNHVYSTOSALSKDPEKRDHNVLLFPTAAGITGMDELYK 238
 Db 181 YVOONTPIIDGPVLLPDPNHVYSTOSALSKDPEKRDHNVLLFPTAAGITGMDELYK 238

RESULT 7
 AAM52335
 ID AAM52335 standard; Protein: 238 AA.

XX AAM52335;
 AC 10-JUL-1998 (first entry)
 DT
 DE Engineered green fluorescent protein.
 XX
 KW Green fluorescent protein; engineered fluorescent protein; mutin;
 KW reporter molecule; immunological assay; protein-protein interaction;
 KW fluorescence resonance energy transfer system; FRET system.
 XX
 OS Synthetic.
 OS Aequorea victoria.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 42
 FT /Label- Cys, Phe, His, Trp, Tyr
 XX
 XX WO9806737-A1.
 XX 19-FEB-1998.
 XX 15-AUG-1997; 97WO-US14593.
 XX 30-AUG-1996; 96US-0706408.
 XX 16-AUG-1996; 96US-0024050.
 XX

PA (AURO-) AURORA BIOSCIENCES.
 PA (REGC.) UNIV CALIFORNIA.
 PA (UOR-) UNIV OREGON.
 XX
 PI Cubitt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
 DR WPI, 1998-159454/14.
 XX
 PT Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants

PS Claim 47; Page -: 120pp; English.

XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52313. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.

XX Sequence 238 AA;

S0

Query Match 99.6%; Score 1265; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6.7e-123;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 MSKGEELFTAVVPLIVELDGVNGHKFSVSGEGSDVYGGKLTITCTTKLTPVWPTL 60
 Db 1 MSKEELFTAVVPLIVELDGVNGHKFSVSGEGSDVYGGKLTITCTTKLTPVWPTL 60

Oy 61 VTFSYGYQCFSRYPDHKKRHDFFKSAPEGYVOORTIFFDDGNYKTRAVEKFGDPLV 120
 Db 61 VTFSYGYQCFSRYPDHKKRHDFFKSAPEGYVOORTIFFDDGNYKTRAVEKFGDPLV 120

Oy 121 NRIELKGIDFEKEDGNILGHKLEYNNSHNYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
 Db 121 NRIELKGIDFEKEDGNILGHKLEYNNSHNYIMADKONGIKVNFKIRHNIEDGSVOLAD 180

Oy 181 YVOONTPIIDGPVLLPDPNHVYSTOSALSKDPEKRDHNVLLFPTAAGITGMDELYK 238
 Db 181 YVOONTPIIDGPVLLPDPNHVYSTOSALSKDPEKRDHNVLLFPTAAGITGMDELYK 238

RESULT 8
 AAM52336
 ID AAM52336 standard; Protein: 238 AA.

XX AAM52336;
 AC 10-JUL-1998 (first entry)
 DT
 DE Engineered green fluorescent protein.
 XX
 KW Green fluorescent protein; engineered fluorescent protein; mutin;
 KW reporter molecule; immunological assay; protein-protein interaction;
 KW fluorescence resonance energy transfer system; FRET system.
 XX
 OS Synthetic.
 OS Aequorea victoria.
 XX

FH Key Location/Qualifiers
 FT Misc-difference 61 /Label- Phe, Tyr, His, Cys
 XX
 XX
 PN M09806737-A1.
 PD 19-FEB-1998.
 XX
 PF 15-AUG-1997; 97WO-US14593.
 XX
 PR 30-AUG-1996; 96US-0706408.
 PR 16-AUG-1996; 96US-0024050.
 XX
 PA (AURO-) AURORA BIOSCIENCES.
 PA (REGC.) UNIV CALIFORNIA.
 PA (UYOR-) UNIV OREGON.
 XX
 PI Cubitt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
 DR WPI; 1998-159454/14.
 XX
 XX Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants
 PS
 PS Claim 47; Page -: 120pp; English.
 XX
 XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52313. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcriptions. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.
 CC
 XX
 SQ Sequence 238 AA;
 Query Match 99.6%; Score 1265; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6.7e-123;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MSKGEELFTAVNPILVELDGDVNGHKFSVSGEGEDVTYGLTKLFTCTTGKLPVMPPTL 60
 DB 1 MSKGEELFTAVNPILVELDGDVNGHKFSVSGEGEDVTYGLTKLFTCTTGKLPVMPPTL 60
 QY 61 VTTFSYGVQCFSRYPDMKRRHDFKFSAMPBGYVOQRTIFFKDGDNKTRAEVFEEDTLV 120
 DB 61 VTTFSYGVQCFSRYPDMKRRHDFKFSAMPBGYVOQRTIFFKDGDNKTRAEVFEEDTLV 120
 QY 121 NRTELKGDIDKEDGNILGHLEKLYNNSHNHYIMADKKNKGIKVFIRHNEDSSVOLAD 180
 DB 121 NRTELKGDIDKEDGNILGHLEKLYNNSHNHYIMADKKNKGIKVFIRHNEDSSVOLAD 180
 QY 181 YVOQNTPIIDGPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEPTAAGITGHMDELK 238
 DB 181 YVOQNTPIIDGPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEPTAAGITGHMDELK 238
 RESULT 9
 AAM52313
 ID AAM52337 standard; Protein: 238 AA.

AC AAM52337;
 XX
 XX 10-JUL-1998 (first entry)
 DT
 XX
 XX Engineered green fluorescent protein.
 DE
 XX
 XX Green fluorescent protein; engineered fluorescent protein; mitein;
 KW reporter molecule; immunological assay; protein-protein interaction;
 KW fluorescence resonance energy transfer system; FRET system.
 XX
 OS Synthetic.
 OS Aequorea victoria.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 62 /Label- Ala, Val, Phe, Ser, Asp, Asn, Gln, Tyr, His, Cys
 FT
 XX
 XX M09806737-A1.
 XX
 XX 19-FEB-1998.
 XX
 PD 15-AUG-1997; 97WO-US14593.
 XX
 PF 30-AUG-1996; 96US-0706408.
 PR 16-AUG-1996; 96US-0024050.
 XX
 XX (AURO-) AURORA BIOSCIENCES.
 PA (REGC.) UNIV CALIFORNIA.
 PA (UYOR-) UNIV OREGON.
 XX
 PI Cubitt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
 DR WPI; 1998-159454/14.
 XX
 XX Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants
 PS
 PS Claim 47; Page -: 120pp; English.
 XX
 XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52313. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcriptions. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.
 CC
 XX
 SQ Sequence 238 AA;
 Query Match 99.6%; Score 1265; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6.7e-123;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MSKGEELFTAVNPILVELDGDVNGHKFSVSGEGEDVTYGLTKLFTCTTGKLPVMPPTL 60
 DB 1 MSKGEELFTAVNPILVELDGDVNGHKFSVSGEGEDVTYGLTKLFTCTTGKLPVMPPTL 60
 QY 61 VTTFSYGVQCFSRYPDMKRRHDFKFSAMPBGYVOQRTIFFKDGDNKTRAEVFEEDTLV 120
 DB 61 VTTFSYGVQCFSRYPDMKRRHDFKFSAMPBGYVOQRTIFFKDGDNKTRAEVFEEDTLV 120

QY 121 NRIELKIDFEKEDNIIIGHKLEIYNNSHNYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
 Db 121 NRIELKIDFEKEDNIIIGHKLEIYNNSHNYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
 QY 181 YYOQNTPIIDGPIVLLPDNHYLSTQSALSCKDNEKRDMHVLLEFYTAAGITGMDLYK 238
 Db 181 YYOQNTPIIDGPIVLLPDNHYLSTQSALSCKDNEKRDMHVLLEFYTAAGITGMDLYK 238

RESULT 10

AAW52338
 ID AAW52338 standard; Protein: 238 AA.

AC AAW52338;
 XX 10-JUL-1998 (first entry)

DE Engineered green fluorescent protein.

KW Green fluorescent protein; engineered fluorescent protein; mutetin;
 KM Reporter molecule; immunological assay; protein-protein interaction;
 XX fluorescence resonance energy transfer system; FRET system.

OS Synthetic.
 XX Aequorea victoria.

FT Key Location/Qualifiers
 FT Misc-difference 68
 FT /label- phe, Tyr, His

PN WO9806737-A1.

PD 19-FEB-1998.

PE 15-AUG-1997; 97WO-US14593.

PR 30-AUG-1996; 96US-0706408.

PR 16-AUG-1996; 96US-0024050.

PA (AURO-) AURORA BIOSCIENCES.

PA (REGC) UNIV CALIFORNIA.

PA (UYOR-) UNIV OREGON.

PI Cubitt AB, Heim R, Ormo MF, Remington JS, Tsien RY;

XX WPI: 1998-159454/14.

DR Nucleic acid encoding mutant green fluorescent proteins having
 XX longer wavelength emission - used as markers for probes and as
 XX components of fluorescent resonant energy transfer systems, also
 XX related vectors and transformants

PS Claim 47; Page -: 120pp; English.

XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAW52333. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.

XX Sequence 238 AA;
 S0

Query Match 99.6%; Score 1265; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6,7e-123;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPIIVELDGVNGHKFSVSGEGDVTYGLKLFICTTGKLPVMPPTL 60
 Db 1 MSKGEELFTAVVPIIVELDGVNGHKFSVSGEGDVTYGLKLFICTTGKLPVMPPTL 60
 QY 61 VTTFSYGQCFSSRPDHHKRRDFKFSAMPEGYVOQRTIFFDDGNKYTRAVEREGDTLV 120
 Db 61 VTTFSYGQCFSSRPDHHKRRDFKFSAMPEGYVOQRTIFFDDGNKYTRAVEREGDTLV 120
 QY 121 NRIELKIDFEKEDNIIIGHKLEIYNNSHNYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
 Db 121 NRIELKIDFEKEDNIIIGHKLEIYNNSHNYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
 QY 181 YYOQNTPIIDGPIVLLPDNHYLSTQSALSCKDNEKRDMHVLLEFYTAAGITGMDLYK 238
 Db 181 YYOQNTPIIDGPIVLLPDNHYLSTQSALSCKDNEKRDMHVLLEFYTAAGITGMDLYK 238

RESULT 11

AAW52344
 ID AAW52344 standard; Protein: 238 AA.

AC AAW52344;
 XX 10-JUL-1998 (first entry)

DE Engineered green fluorescent protein.

KW Green fluorescent protein; engineered fluorescent protein; mutetin;
 KM Reporter molecule; immunological assay; protein-protein interaction;
 XX fluorescence resonance energy transfer system; FRET system.

OS Synthetic.
 XX Aequorea victoria.

FT Key Location/Qualifiers
 FT Misc-difference 150
 FT /label- phe, Tyr, His

PN WO9806737-A1.

PD 19-FEB-1998.

PE 15-AUG-1997; 97WO-US14593.

PR 30-AUG-1996; 96US-0706408.

PR 16-AUG-1996; 96US-0024050.

PA (AURO-) AURORA BIOSCIENCES.

PA (REGC) UNIV CALIFORNIA.

PA (UYOR-) UNIV OREGON.

PI Cubitt AB, Heim R, Ormo MF, Remington JS, Tsien RY;

XX WPI: 1998-159454/14.

DR Nucleic acid encoding mutant green fluorescent proteins having
 XX longer wavelength emission - used as markers for probes and as
 XX components of fluorescent resonant energy transfer systems, also
 XX related vectors and transformants

PS Claim 47; Page -: 120pp; English.

XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAW52333. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.

Y, W or F) and having fluorescent properties different from GFP. (II) CC are useful as reporter molecules in immunological or hybridisation CC assays, for monitoring proteins in cells and detecting induction of CC transcription. They are also useful as components of FRET (fluorescence resonance energy transfer) systems, e.g. for detecting protein-protein CC interactions, cleavage of substrates and changes in potential across a CC membrane, and further for making fluorescent substrates for protein CC kinase. (II) are easily distinguished from known green and blue CC fluorescent proteins, so allow simultaneous measurements of at least 2 CC proteins in a cell. They emit, and are excited, at relatively long CC wavelengths where phototoxicity and auto-fluorescence are reduced.

Sequence 238 AA;

Query Match 99.6%; Score 1265; DB 19; Length 238;
Best Local Similarity 99.6%; Pred. No. 6,7e-123;
Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVNPILVELDGVNGHKFSVSGEGSDVYTKLTKICTTGKLPVMPPTL 60
DB 1 MSKGEELFTAVNPILVELDGVNGHKFSVSGEGSDVYTKLTKICTTGKLPVMPPTL 60
QY 61 VTFPSYGVQCFSRPDMKRRHDFPKSAMPEGYVOQRTIFPKDGNKTRAEVKEGDTLV 120
DB 61 VTFPSYGVQCFSRPDMKRRHDFPKSAMPEGYVOQRTIFPKDGNKTRAEVKEGDTLV 120
QY 121 NRIELKIDFKEDGNILGHKLEYNNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
DB 121 NRIELKIDFKEDGNILGHKLEYNNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 181 YVOQNPILDGPVLLPNNHYLSTOSALSCKPNEKRDMVLEFVTAAGITHGMDLYK 238
DB 181 YVOQNPILDGPVLLPNNHYLSTOSALSCKPNEKRDMVLEFVTAAGITHGMDLYK 238

RESULT 12

AAW52346
ID AAW52346 standard; Protein: 238 AA.

AC AAW52346;

DT 10-JUL-1998 (first entry)

DE Engineered green fluorescent protein.

KW Green fluorescent protein; engineered fluorescent protein; muten;

KW reporter molecule; immunological assay; protein-protein interaction;

KW fluorescence resonance energy transfer system; FRET system.

OS Synthetic.

OS Aequorea victoria.

FT Key Location/Qualifiers

FT MISC-difference 167 /Label= Phe, Tyr, His

PN W09806737-A1.

PD 19-FEB-1998.

PE 15-AUG-1997; 97WO-US14593.

PR 30-AUG-1996; 96US-0706408.

PR 16-AUG-1996; 96US-0024050.

PA (AURO-) AURORA BIOSCIENCES.

PA (REGC) UNIV CALIFORNIA.

PA (UYOR-) UNIV OREGON.

PI Cubitt AB, Heim R, Ormo MF, Remington JS, Tsien RY;

XX WPI; 1998-159454/14.

PT Nucleic acid encoding mutant green fluorescent proteins having
PT longer wavelength emission - used as markers for probes and as
PT components of fluorescent resonant energy transfer systems, also
PT related vectors and transformants

PS Claim 47: Page -: 120pp; English.

XX This sequence is an engineered Aequorea victoria green fluorescent
CC protein (GFP) of the invention, created from the sequence shown in
CC AAW52313. The GFP DNA was mutated to produce the nucleic acids (I) of
CC the invention, which encode functional, engineered fluorescent proteins
CC (II) having largely the same sequence as Aequorea green fluorescent
CC protein (GFP) but differing by at least the substitution T203X (X = H,
CC Y, W or F) and having fluorescent properties different from GFP. (II)
CC are useful as reporter molecules in immunological or hybridisation
CC assays, for monitoring proteins in cells and detecting induction of
CC transcription. They are also useful as components of FRET (fluorescence
CC resonance energy transfer) systems, e.g. for detecting protein-protein
CC interactions, cleavage of substrates and changes in potential across a
CC membrane, and further for making fluorescent substrates for protein
CC kinase. (II) are easily distinguished from known green and blue
CC fluorescent proteins, so allow simultaneous measurements of at least 2
CC proteins in a cell. They emit, and are excited, at relatively long
CC wavelengths where phototoxicity and auto-fluorescence are reduced.

Sequence 238 AA;

Query Match 99.6%; Score 1265; DB 19; Length 238;
Best Local Similarity 99.6%; Pred. No. 6,7e-123;
Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVNPILVELDGVNGHKFSVSGEGSDVYTKLTKICTTGKLPVMPPTL 60
DB 1 MSKGEELFTAVNPILVELDGVNGHKFSVSGEGSDVYTKLTKICTTGKLPVMPPTL 60
QY 61 VTFPSYGVQCFSRPDMKRRHDFPKSAMPEGYVOQRTIFPKDGNKTRAEVKEGDTLV 120
DB 61 VTFPSYGVQCFSRPDMKRRHDFPKSAMPEGYVOQRTIFPKDGNKTRAEVKEGDTLV 120
QY 121 NRIELKIDFKEDGNILGHKLEYNNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
DB 121 NRIELKIDFKEDGNILGHKLEYNNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 181 YVOQNPILDGPVLLPNNHYLSTOSALSCKPNEKRDMVLEFVTAAGITHGMDLYK 238
DB 181 YVOQNPILDGPVLLPNNHYLSTOSALSCKPNEKRDMVLEFVTAAGITHGMDLYK 238

RESULT 13

AAW52349
ID AAW52349 standard; Protein: 238 AA.

AC AAW52349;

DT 10-JUL-1998 (first entry)

DE Engineered green fluorescent protein.

KW Green fluorescent protein; engineered fluorescent protein; muten;

KW reporter molecule; immunological assay; protein-protein interaction;

KW fluorescence resonance energy transfer system; FRET system.

OS Synthetic.

OS Aequorea victoria.

FT Key Location/Qualifiers

FT MISC-difference 220 /Label= His, Asn, Gln, Thr

PN W09806737-A1.

PD 19-FEB-1998.

RESULT 15
AAW52331
ID AAW52331 standard; Protein: 238 AA.
XX
AC AAW52331;
XX
DT 10-JUL-1998 (first entry)
XX
DE Engineered green fluorescent protein S65C.
XX
KW Green fluorescent protein; engineered fluorescent protein; mutain;
KW reporter molecule; immunological assay; protein-protein interaction;
KW fluorescence resonance energy transfer system; FRET system.
XX
OS Synthetic.
OS Aequorea victoria.
XX
FH Key Location/Qualifiers
FT Misc-difference 65 /label= S65C
FT /note= "Ser to Cys mutation"
XX
XX
PN MO9806737-A1.
XX
XX 19-FEB-1998.
XX
PF 15-AUG-1997; 97WO-US14593.
XX
PR 30-AUG-1996; 96US-0706408.
PR 16-AUG-1996; 96US-0024050.
XX
XX (AURO-) AUROREA BIOSCIENCES.
PA (REGC) UNIV CALIFORNIA.
PA (UYOR-) UNIV OREGON.
XX
PI Cuditt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
XX
XX WPI: 1998-159454/14.
XX
XX Nucleic acid encoding mutant green fluorescent proteins having
PT longer wavelength emission - used as markers for probes and as
PT components of fluorescent resonant energy transfer systems, also
PT related vectors and transforants
XX
XX
PS Claim 5; Page -: 120pp; English.
XX
XX This sequence is an engineered Aequorea victoria green fluorescent
CC protein (GFP) of the invention, created from the sequence shown in
CC AAW52313. The GFP DNA was mutated to produce the nucleic acids (I) of
CC the invention, which encode functional, engineered fluorescent proteins
CC (II) having largely the same sequence as Aequorea green fluorescent
CC protein (GFP) but differing by at least the substitution T203X (X = H,
CC Y, W or F) and having fluorescent properties different from GFP. (II)
CC are useful as reporter molecules in immunological or hybridisation
CC assays, for monitoring proteins in cells and detecting induction of
CC transcription. They are also useful as components of FRET (fluorescence
CC resonance energy transfer) systems, e.g. for detecting protein-protein
CC interactions, cleavage of substrates and changes in potential across a
CC membrane, and further for making fluorescent substrates for protein
CC kinase. (II) are easily distinguished from known green and blue
CC fluorescent proteins, so allow simultaneous measurements of at least 2
CC proteins in a cell. They emit, and are excited, at relatively long
CC wavelengths where phototoxicity and auto-fluorescence are reduced.
XX
SO Sequence 238 AA;

Query Match 99.6%; Score 1265; DB 19; Length 238;
Best Local Similarity 99.6%; Pred. No. 6,7e-123;
Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 MSKGEEFTAVVPILVELDGDVNGHKFSVSGEGEDVTYGKLTLEKICTTGKLPVMPPTL 60

Db 1 MSKGEEFTAVVPILVELDGDVNGHKFSVSGEGEDVTYGKLTLEKICTTGKLPVMPPTL 60
QY 61 VTTFSGVQCPSPRYPDHMKRRHDFKSAPEGYVOORTLFFDDGNYKTRAEVKEGDTIV 120
Db 61 VTTFSGVQCPSPRYPDHMKRRHDFKSAPEGYVOORTLFFDDGNYKTRAEVKEGDTIV 120
QY 121 NRILKGIIDFKEDGNILGHLKLEYNNSHNYIMADKQKNGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRILKGIIDFKEDGNILGHLKLEYNNSHNYIMADKQKNGIKVFKIRHNIEDGSVOLAD 180
QY 181 YVOQNTPILOGPVLLPDNHYLSTOSALSADPNKRDHMYLLEFVTAAGITGMDLYK 238
Db 181 YVOQNTPILOGPVLLPDNHYLSTOSALSADPNKRDHMYLLEFVTAAGITGMDLYK 238

Search completed: July 11, 2003, 17:56:46
Job time : 74 secs